3 確率

3.1 基本的な概念

- 試行：実験や観測などを試行 (trial) と呼ぶ。
- 標本空間 Ω：試行の結果の集合を標本空間 (sample space) と言う。
- 標本点 ω：標本空間 Ω の各要素を標本点 (sample point) と言う。

例 3.1. (コイン投げ) コインを 1 回投げ、表か裏かを観測するという試行を考える。このとき標本空間は

$$\Omega = \{\,裏, 表\,\}\ \Big(= \{0, 1\}\Big)$$

で与えられる。"裏" $(= \omega)$ は 1 つの標本点である。//

例 3.2. (サイコロ投げ) サイコロを 1 回投げ、出目を観測するという試行を考える。このとき標本空間は

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

となる。例えば、"2" は 1 つの標本点。//

標本空間 Ω の部分集合を事象 (event) と言う。

- 標本空間 Ω は Ω の部分集合であるから、Ω は 1 つの事象である (全事象)。
- 空集合 $\phi = \{\}$ も Ω の部分集合であるから、ϕ は 1 つの事象である (空事象)。
- ただ 1 つの標本点のみからなる事象 $\{\omega\}$ を根元事象という。

例 3.3. (コイン投げ) コインを 3 回投げて、表 $(= 1)$ が出るか裏 $(= 0)$ が出るかを観測するという試行を考えると、標本空間 Ω は

$$\Omega = \Big\{(0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,1),\ (1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)\Big\}$$

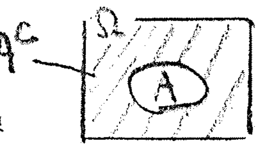
となる ($2^3 = 8$ 個の標本点)。事象の例を幾つか挙げる。

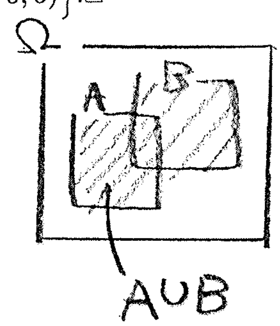
$$A = \Big\{1\,回目は表\Big\} = \Big\{(1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)\Big\} \tag{3.1}$$

$$B = \Big\{ちょうど\,2\,回表が出る\Big\} = \Big\{(0,1,1),\ (1,0,1),\ (1,1,0)\Big\} \tag{3.2}$$

$$C = \Big\{ちょうど\,1\,回表が出る\,\Big\} = \Big\{(0,0,1),\ (0,1,0),\ (1,0,0)\Big\}.\square \tag{3.3}$$

3.2 事象の演算など





- 補事象、余事象 ... A が起こらないという事象

$$A^c = A\,の補事象 = \{A\,に含まれない標本点\,\} \tag{3.4}$$

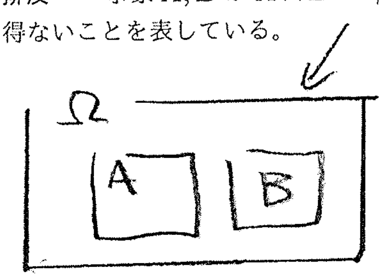
- 和事象 ... A または B が起こるという事象

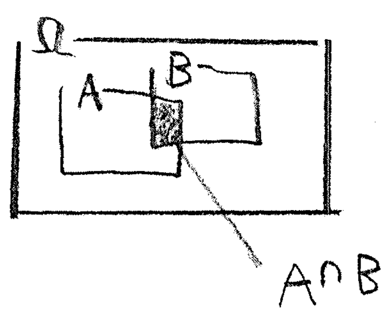
$$A \cup B = A\,と\,B\,の和事象 = \{A\,または\,B\,に含まれる標本点\,\} \tag{3.5}$$

- 積事象 ... A と B の両方が起こるという事象

$$A \cap B = A\,と\,B\,の積事象 = \{A\,と\,B\,の両方に含まれる標本点\,\} \tag{3.6}$$

- 互いに排反 ... 事象 A, B が $A \cap B = \phi$ を満たすとき、A と B は互いに排反であると言う。A と B は同時に起こり得ないことを表している。





例 3.4. A, B, C を3つの事象とする。次の各場合を表す式を書け。

1. A だけが起こる;
2. A と B が起こるが C は起こらない;
3. 少なくとも1つが起こる;
4. 少なくとも2つが起こる;
5. 1つだけが起こる;
6. 2つ起こるが3つともは起こらない。

答) 1. $A \cap B^c \cap C^c$、2. $A \cap B \cap C^c$、3. $A \cup B \cup C$, 4. $(A \cap B) \cup (A \cap C) \cup (B \cap C)$、5. $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$、6. $(A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C)$.

3.3 確率の基本公式

公式 3.1. (1) A と B が互いに排反ならば、$P(A \cup B) = P(A) + P(B)$.

(2) $A_1, A_2, \ldots, A_n$ が互いに排反ならば、$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$.

(3) $P(A^c) = 1 - P(A)$.

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

公式 3.2. (基本公式，各根元事象が等確率の場合) 標本空間が有限集合 $\Omega = \{x_1, x_2, \cdots, x_N\}$ でかつ，各根元事象の確率が等しい場合，すなわち

$$P(\{x_1\}) = P(\{x_2\}) = \cdots = P(\{x_N\}) = \frac{1}{N} \tag{3.7}$$

となる場合，任意の事象 A の確率は

$$P(A) = \frac{A\text{に含まれる標本点の数}}{N} \tag{3.8}$$

となる。□

公式 3.3. (コイン投げ) 歪みのないコインを n 回投げたとき表が k 回出る事象の確率は

$$P\Big(\{k\text{ 回表が出る }\}\Big) = \frac{{}_nC_k}{2^n} \quad (k = 0, 1, \cdots, n).$$

である[1]。これは次のようにして納得できる。

$n = 5$、$k = 2$ として説明する（理解できればすぐに一般化できる）。コインを5回投げて表 (=1) か裏 (=0) かを観測する試行の標本空間は

$$\Omega = \{(0,0,0,0,0),\ (0,0,0,0,1),\ \cdots,\ (1,1,1,1,1)\} = \{(a_1, a_2, a_3, a_4, a_5) \mid a_i\text{は 0 または 1 } (i = 1,2,3,4,5)\}$$

となる。これは $2^5 = 32$ 個の標本点からなる。一つ一つの確率（根元事象の確率）は全て等しく 1/32 である。従って、{2 回表が出る } という事象の確率はこの事象に含まれる標本点を数えればよい。

$$\{2\text{ 回表が出る }\} = \{(0,0,0,1,1),\ (0,0,1,0,1),\ \cdots,\ (1,1,0,0,0)\}$$

上の事象に含まれる標本点の数は ${}_5C_2 = 10$ である（「5個の異なる場所があり、そこから2つ選んで1を入れ、残った場所に0を入れる」と考える）から、

$$P\Big(\{2\text{ 回表が出る }\}\Big) = \frac{{}_5C_2}{2^5} = \frac{10}{32} = \frac{5}{16}$$

である。□

[1] ここで、${}_nC_k$ は n 個の異なるものの中から k 個を選ぶ選び方の数（組み合わせの数）であり、${}_nC_k = \frac{n!}{k!(n-k)!}$。例えば、A,B,C,D,E 君の5人の中から生徒会役員2名を選ぶ選び方は、${}_5C_2 = 5!/(2!3!) = 10$ 通り。念のため確認しておくと、階乗の定義は $k! = k \times (k-1) \times \cdots \times 2 \times 1$。よって $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$。なお 0! と約束することも思い出そう。

3.4　条件付確率

応用においては、確率は「割合」という意味に多く用いられるので注意する。この場合、条件付確率 $P(A|B) = P(A\cap B)/P(B)$ は「B に占める $A\cap B$ の割合」という意味となる。

例えば、表 1 において

$$P(A\cap B) = \text{「40 歳未満」でかつ、「そう思う（生死は本人の判断に任せるべき）」人の割合} = \frac{217}{2017} = 0.108$$
$$P(A|B) = \text{「40 歳未満」に占める、「そう思う」の割合} = \frac{217}{496} = 0.438$$

となる。また、表 2 のように条件付確率の値のみが与えられている場合も多い。次の例のように枝分かれの構造（図 3 参照）がある場合も条件付確率の値のみが与えられている例である。

例 3.5. (くじ)100 本中 20 本に「当たり」のあるくじを次々と引く。$\{i$ 回目に「当たり」を引く $\}$ という事象を A_i とおく。例えば、

$$A_1 = \{1\text{ 回目に当たりを引く}\},\ A_1\cap A_2^c = \{1\text{ 回目当たり、}2\text{ 回目はずれ}\}$$

などである（もちろn $P(A_1) = 20/100$、$P(A_1^c) = 80/100$）。この問題においては、条件付確率

$$P(A_2|A_1) = \frac{19}{99},\ P(A_3|A_1\cap A_2) = \frac{18}{98},\ P(A_4^c|A_1\cap A_2\cap A_3) = \frac{80}{97}$$

などの値は直ぐに分かる。(因みに $P(A_1) = P(A_2) = \cdots = P(A_{100}) = 20/100$。) □

次は、定義に基づいて条件付確率の計算をしないとその値が分からない例。

例 3.6. (条件付確率の練習, 人によっては意外に感じる例) 子供 2 人の家庭は、上が男で下が女という家庭、その逆の家庭、2 人とも男、2 人とも女という 4 通りがあるが、それらは各々1/4 の確率で分布しているとする。今、ある家庭を訪問したところ子供が 2 人であること、内 1 人が男であることが分かった。もう 1 人も男である確率を求めよ。

(解) 性別を上の子、下の子の順に表すならば、標本空間は

$$\Omega = \{\text{男女, 女男, 男男, 女女}\} \tag{3.9}$$

求める確率は、$P(A|B)$ である。但し、

$$B = \{1\text{ 人が男}\} = \{\text{男女, 女男, 男男}\},\quad A = \{2\text{ 人が男}\} = \{\text{男男}\} \tag{3.10}$$

である。従って、条件付確率の定義と $A\cap B = \{\text{男男}\}$ を使うと、

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(\{\text{男男}\})}{P(\{\text{男女, 女男, 男男}\})} = \frac{1/4}{3/4} = \frac{1}{3}. \tag{3.11}$$

となる. □

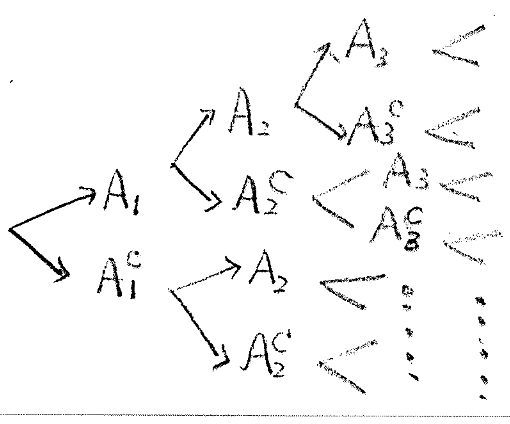
表1

	そう思う A	そう思わない Ac	計
40歳未満　B	217	279	496
40歳以上　Bc	440	1081	1521
計	657	1360	2017

表2

	そう思う A	そう思わない Ac	計
40歳未満　B	0.438	0.562	1.00
40歳以上　Bc	0.289	0.711	1.00

図1



3.5 全確率の公式とベイズの定理

例 3.7. (練習) ある病院の通院患者の血圧について次のことが知られている：

$$\text{正常}\ (H_1) : \text{軽度高血圧}\ (H_2) : \text{重度高血圧}\ (H_3) = 50 : 30 : 20$$

そのそれぞれが、症状 A を示す割合は、正常者の 10%、軽度高血圧者の 20%、重度高血圧者の 50%であるとする。通院患者を 1 人任意に選んだ時、その人が症状 A を示す確率は幾らか。

(解) 標本空間を次のように分割すると、

$$\Omega = H_1 \cup H_2 \cup H_3 \tag{3.12}$$

題意より、

$$P(H_1) = 0.5,\ P(H_2) = 0.3,\ P(H_3) = 0.2. \tag{3.13}$$

一方、症状 A を示すという事象を A と表せば、

$$P(A|H_1) = 0.1,\ P(A|H_2) = 0.2,\ P(A|H_3) = 0.5 \tag{3.14}$$

であるから、全確率の公式より、

$$\begin{aligned} P(A) &= P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3) \\ &= 0.1 \times 0.5 + 0.2 \times 0.3 + 0.5 \times 0.2 \\ &= 0.05 + 0.06 + 0.10 = 0.21. \end{aligned} \tag{3.15}$$

よって、$P(A) = 0.21$。

また、症状 A を示した人が重度高血圧者である確率は、ベイズの定理より

$$P(H_3|A) = \frac{P(A \cap H_3)}{P(A)} = \frac{P(A|H_3)P(H_3)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3)} \tag{3.16}$$

$$= \frac{0.10}{0.05 + 0.06 + 0.10} = 0.476 \tag{3.17}$$

であり、約 48 %である。□

例 3.8. (検査の精度) ある病気の 1 次検査に用いられる方法では、罹患している人の 90%に対して陽性反応を示すが、罹患していない人の 5%にも陽性反応を示す。これまでの統計から、1 次検査を受ける人のおよそ 2%がその病気に罹患していることが知られているとする。新たに 1 次検査を受けた人が陽性反応を示した。その人が真に罹患している確率を求めよ。

例 3.9. (択一問題に正解した人が本当に正解を知っている確率) 面接担当者が受験者に択一問題を出す。受験者が本当に正解を知っているかどうかに関心があるものとする。4 人に 1 人が正解を知っているという程度の難しさの問題を出題するものとする (**正解率 $p = 1/4$**)。この問題を 3 択問題 (**$m = 3$ と置く**) で出題するものとする。

$$A = \{\ \text{受験者は本当に正解を知っている}\ \},\ \ B = \{\ \text{受験者は正解する}\ \} \tag{3.18}$$

事象 B にはまぐれ当りも含まれる。

(1) $P(A)$、$P(B|A)$ および $P(B|A^c)$ は幾らか。 (2) 解答者が正解したという条件の下で、その人がまぐれではなく本当に正解を知っている確率を求めよ。(3) 100 人に 1 人しか正解を知らないという難しい問題を 100 択問題で出題したとする。解答者が正解したという条件の下で、その人がまぐれではなく本当に正解を知っている確率を求めよ。

(解)(1) 4 人に 1 人が正答するのだから、$P(A) = 1/4$。また、正解を知っている人は必ず正答するから $P(B|A) = 1$。正解を知らない人はまぐれ当たりしかないから $P(B|A^c) = 1/3$。(2) 乗法公式と全確率公式により $P(A \cap B) = P(B|A)P(A) = 1 \times (1/4) = 1/4$ であり, $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) = 1 \times (1/4) + (1/3) \times (3/4) = 1/2$ となるから,$P(A|B) = P(A \cap B)/P(B) = (1/4)/(1/2) = 1/2$。(3) $p = 1/100$、$m = 100$ として、上の展開を辿ると $P(A|B) = (1/100)/\{1/100 + (1/100) \times (99/100)\} = 100/199 \approx 1/2$ を得る。