

確率と統計

いろはす

2023 年 11 月 20 日

第 1 章

確率論

1.1 確率空間と確率関数

確率を考えるときには、次の 2 つが定まっていることが前提となる:

1. **確率空間** S : 起こりうること (事象) を考えるための土台となる集合.
2. **確率関数** P : 各事象 (S の部分集合^{*1}) に対し 確率 (0 以上 1 以下の値) を割り当てる規則.

より正確に言うと、 P は確率空間 S **上**の確率関数と呼ばれる. また、集合 S と規則 P の組 (S, P) をもって確率空間と呼ぶこともある.

事象 A に対して、その確率を $P(A)$ と書く.

コメント 1.1.1

事象を表わすのには、主に A, B, C, \dots などの記号を使う.

空集合 \emptyset もまた一つの事象 (**空事象**という) として扱う^{*2}. これは言わば、「絶対に起こらない事象」を表わす.

また、確率空間 S 自身も (S の部分集合だから) 事象であり、**全事象**などと呼ばれる. これは「確実に起こる事象」を表わす.

^{*1} 厳密には、 S の部分集合全部が事象になるわけではない. これは、確率とか考えられないような変ちくりんな集合もこの世には存在するためであるが、この辺を深堀りし出すとかなり高度な話になってしまうのでここでは気にしないことにする.

^{*2} 集合論では、空集合は全ての集合の部分集合と見做される.

例 1.1.1: サイコロ投げ

サイコロ投げでは、(フツーは^a) 次のように定める：

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

$$P(A) = \frac{\#A}{6} \quad (\#A \text{ は集合 } A \text{ の要素数}).$$

S の一つ一つの要素が、サイコロを投げたときに「出る目」を表わす。事象は S の部分集合だから、例えば

- $A = \{3\}$ (3 の目が出る事象)
- $B = \{2, 4, 6\}$ (偶数の目が出る事象)
- $C = \{5, 6\}$ (5 以上の目が出る事象)

などが該当する。これらの確率は、それぞれ次の通りである：

$$P(A) = \frac{\#A}{6} = \frac{1}{6},$$

$$P(B) = \frac{\#B}{6} = \frac{3}{6} = \frac{1}{2},$$

$$P(C) = \frac{\#C}{6} = \frac{2}{6} = \frac{1}{3}.$$

^a 後ほど述べるように、確率関数 P の定め方は一通りではない。後述の「ルール」を満たしていれば、数学的にはどんなものも許される。規則 P としてどんなものを採用するかは、個々の状況に合わせて人間が決めることである。

例 1.1.2: ダーツ

ダーツを的に向かって一回だけ投げる行為を考える（ただし、的に当たらなかった場合はやり直すことにする）。この場合、ダーツの的の面全体を S と考えることができる。事象は、 S の一部分を切り抜いた領域の一つ一つである。 A という領域は、意味として「ダーツが A 内のどこかに当たる事象」を表わすのである。

確率関数の決め方は、この場合色々ある。投げる人とか、色々な条件で変わってくる。その時々でふさわしいものを使うべきである。

コメント 1.1.2

数学（の確率を扱う分野＝確率論）は、単なる計算の「枠組み」に過ぎないので、確率関数の「決め方」までは教えてくれない。それは個々の状況に応じて人間が適切に決めることである。

高校数学だと、よく

$$P(A) = \frac{A \text{ が起こる場合の数}}{\text{起こりうる事象のすべての場合の数}} \quad (1.1.1)$$

みたいな説明を受ける。我々の記号で書くと、これは

$$P(A) = \frac{\#A}{\#S} \quad (1.1.2)$$

と同じ意味である。

しかし、これは「確率の決め方」の一つに過ぎない。これ以外にも色々な決め方があるだろうし、場合によっては他の「決め方」の方が現実をうまく表現できることもある。例えば、例 1.1.1 の P は式 (1.1.1) の「決め方」を用いた例である。しかし、サイコロ投げにおいてこの P を使うことが、常に「妥当な」選択だと言い切れるだろうか？もしサイコロが、イカサマ師が使うような、重心がひどく偏っていてやたら 1 の目ばかり出るようなサイコロだったらどうするのだろうか？そのような場合、明らかに例 1.1.1 の P を用いるのは不適切であろう。

上のコメントで述べた通り、確率の「決め方」は一通りではない。しかし、その「決め方」にも最低限のルールがある。確率空間 S 上の確率関数 P は、必ず以下のルールを満たさなくてはならない：

- ルール①：どんな事象 A に対しても、 $P(A) \geq 0$.
- ルール②：どんな事象 A, B に対しても、

$$A \cap B = \emptyset \text{ ならば } P(A \cup B) = P(A) + P(B).$$

- ルール③： $P(S) = 1$.

上のルールを満たすような P のみが、 S 上の確率関数と呼ばれるのである。なお、例 1.1.1 の P がこれらのルールを満たしていることは容易に確かめられる。

定理 1.1.1: 確率の基本性質

確率空間 S 上の確率関数 P に対して、次が成り立つ：

1. $P(\emptyset) = 0$.

2. どんな事象 A, B に対しても,

$$A \subset B \text{ ならば } P(A) \leq P(B).$$

3. どんな事象 A に対しても, $0 \leq P(A) \leq 1$.

証明.

1. $\emptyset \cap \emptyset = \emptyset \cup \emptyset = \emptyset$ だから, ルール②より

$$P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset).$$

両辺から $P(\emptyset)$ を引けば $P(\emptyset) = 0$ を得る.

2. $A' = B \setminus A$ (記号 \setminus は**差集合**の記号で, $B \setminus A$ は「 B に属しかつ A に属さないものの集合」を表わす) とおく. このとき, $A \cap A' = \emptyset$ かつ, (仮定 $A \subset B$ より) $B = A \cup A'$ だから, ルール②より

$$P(B) = P(A \cup A') = P(A) + P(A'). \quad (1.1.3)$$

一方, ルール①より $P(A') \geq 0$ であるから,

$$P(A) + P(A') \geq P(A). \quad (1.1.4)$$

式 (1.1.3), (1.1.4) より $P(B) \geq P(A)$.

3. 事象は S の部分集合なので, $A \subset S$. これと 2. より

$$P(A) \leq P(S).$$

これとルール③ ($P(S) = 1$) より

$$P(A) \leq 1. \quad (1.1.5)$$

また, ルール①より

$$0 \leq P(A). \quad (1.1.6)$$

式 (1.1.5), (1.1.6) より $0 \leq P(A) \leq 1$. ■

1.2 条件付き確率

平成 30 年の国民健康・栄養調査では、糖尿病患者の割合は男性が 18.7%, 女性が 9.3% である。割合と確率は厳密に言うとは別物であるが、今これらの数値を確率と解釈することになると、これらの確率には「男性」や「女性」といった「条件」が付いている。

確率論では、このような確率を**条件付き確率**といい、 $P(\text{事象} \mid \text{条件})$ という形式で書く。上の例で言えば、

$$P(\text{糖尿病} \mid \text{男性}) = 0.187, P(\text{糖尿病} \mid \text{女性}) = 0.093$$

のようになる。

これらは、**同時確率** $P(\text{糖尿病} \cap \text{男性})$, $P(\text{糖尿病} \cap \text{女性})$ とは全くの**別物**であることに注意しなければならない。実際、仮に男性と女性の割合が正確に 0.5 ずつだったとすると、

$$\begin{aligned} P(\text{糖尿病} \cap \text{男性}) &= P(\text{糖尿病} \mid \text{男性})P(\text{男性}) = 0.187 \times 0.5 = 0.0935, \\ P(\text{糖尿病} \cap \text{女性}) &= P(\text{糖尿病} \mid \text{女性})P(\text{女性}) = 0.093 \times 0.5 = 0.0465 \end{aligned}$$

となる。この例を見ると、

$$P(\text{事象} \mid \text{条件}) = \frac{P(\text{事象} \cap \text{条件})}{P(\text{条件})}$$

という関係が成り立っているのがわかる。

以上の例を踏まえて、条件付き確率の定義を明確にしよう。確率空間 S 上の確率関数 P を固定する。確率が 0 でない事象 C （これが条件 (Condition) を表す）に対して、 S 上の確率関数 P_C を

$$P_C(A) = \frac{P(A \cap C)}{P(C)} \quad (1.2.1)$$

にて定義する（これが先程の「ルール」を満たしていることは容易に確認できる）。この P_C を、条件 C のもとでの**条件付き確率**という。なお、事象 A に対して $P_C(A)$ を $P(A|C)$ と表記する。

例 1.2.1

上の糖尿病の例で、一人の国民を無作為に抽出したとき、その人（X さんとする）が糖尿病だったとする。このとき、X さんが女性である確率を求めてみよう。確率論の記号で書けば、求めるものは

$$P(\text{女性} \mid \text{糖尿病})$$

である。条件付き確率の定義に従えば、これは次の式で求められる：

$$P(\text{女性} \mid \text{糖尿病}) = \frac{P(\text{女性} \cap \text{糖尿病})}{P(\text{糖尿病})}. \quad (1.2.2)$$

ここで、分子の $P(\text{女性} \cap \text{糖尿病})$ は上で既に見た通りである：

$$P(\text{女性} \cap \text{糖尿病}) = 0.0465.$$

分母は次のようにして求められる。まず、

$$\begin{aligned} \text{糖尿病} &= (\text{糖尿病} \cap \text{男性}) \cup (\text{糖尿病} \cap \text{女性}), \\ (\text{糖尿病} \cap \text{男性}) \cap (\text{糖尿病} \cap \text{女性}) &= \emptyset \end{aligned}$$

であるから、ルール②より

$$\begin{aligned} P(\text{糖尿病}) &= P((\text{糖尿病} \cap \text{男性}) \cup (\text{糖尿病} \cap \text{女性})) \\ &= P(\text{糖尿病} \cap \text{男性}) + P(\text{糖尿病} \cap \text{女性}) \\ &= 0.0935 + 0.0465 \\ &= 0.14. \end{aligned}$$

以上を式 (1.2.2) に代入し、答えは

$$P(\text{女性} \mid \text{糖尿病}) = \frac{0.0465}{0.14} \simeq 0.332$$

と求まる。

1.3 確率変数

確率変数とは、確率的に変動する量を概念化したものである。例えば、サイコロ投げなら出る目の数値を X とおくと、これは確率変数になる。ダーツの場合なら、ダーツが当たった場所の x 座標と y 座標をそれぞれ X, Y とおくと、これらは確率変数になる。一般に、確率変数には X, Y, Z などの文字を用いるのが通例である。

確率変数の数学的な定義は、次のようになる。再び、確率空間 S とその上の確率関数 P を固定する。ここで S 上の**確率変数**とは、 S の要素に一つの実数に対応させる規則（すなわち S 上の**関数**）である。

コメント 1.3.1

確率**変数**と名が付くものの実体が関数であるというのは、一見して変な感じがする。しかし、上のサイコロやダーツの例を省みても、 S の一つ一つの要素に対して、変数の値が一つ確定するという構造になっている。従って「確率変数とは何か？」と問うたとき、この対応関係そのものをもって定義とするのが最もスッキリ行くのである。

確率変数 X が取り得る値を、 X の実現値という。もし X が有限個の値しか取り得なければ、それらを x_1, \dots, x_n のような記号で表わすのが通例である。

a を S の要素とする。確率変数 X によって a が対応させられる値を $X(a)$ と書く。

例 1.3.1: 確率変数の例：サイコロ投げ

サイコロ投げで出る目を X とすると、 X の実現値は $1, \dots, 6$ である。 a を S の要素とすると、 X は規則

$$X(a) = a$$

で表わされる。

実数値関数（実数を実数に対応させる規則） f があれば、これを使って確率変数

X から新たな確率変数 $f(X)$ を作ることができる。つまり、

$$f(X)(a) = f(X(a))$$

と定義するのである。なお、ここでは f を 1 変数の関数としたが、2 変数やより沢山の変数をもつ関数でも同様である。

例 1.3.2: 確率変数の例：ダーツ

ダーツが当たった場所の x 座標と y 座標をそれぞれ X, Y とする。なお、的の中心の座標を原点 $(0, 0)$ とする。このとき、座標 (x, y) の原点からの距離は

$$r(x, y) = \sqrt{x^2 + y^2}$$

で与えられる。そこで、新たな確率変数 R を

$$R = r(X, Y) = \sqrt{X^2 + Y^2}$$

にて定義すると、これはダーツが当たった場所の原点からの距離を表わす。

確率変数 X と、実数 t や、実数の部分集合 A に対して、

$$X = t, X < t, X \in A$$

などは事象である。これらをより「きちんと」書くと、

$$\{a \in S \mid X(a) = t\}, \{a \in S \mid X(a) < t\}, \{a \in S \mid X(a) \in A\}$$

となる ($\{a \in S \mid \text{条件}\}$ で、「条件を満たす S の要素 a の集合」を表わす)。確かに、これらは S の部分集合であるから、事象である。

より一般に、実数についての条件 $\varphi(x)$ に対して、 $\varphi(X)$ は事象である。これもより「きちんと」書けば、

$$\{a \in S \mid \varphi(X(a))\}$$

となる。

$\varphi(X)$ は事象であるから、その確率が定義できる。これを、

$$P(\varphi(X))$$

のように書く。上の具体例で書くなら、これは

$$P(X = t), P(X < t), P(X \in A)$$

などにあたる．

例 1.3.3: 確率変数の例：サイコロ投げ 2

再び，例 1.3.1 のように確率変数 X を定める．このとき，

$$\begin{aligned} P(X = 1) &= \cdots = P(X = 6) = \frac{1}{6}, \\ P(X > 4) &= \frac{1}{3}, \\ P(X \text{ は偶数}) &= \frac{1}{2}. \end{aligned}$$

となる．

1.4 確率分布（実現値有限個の場合）

再び，確率空間 S とその上の確率関数 P を固定する．

ここでは， S 上の確率変数 X が有限個の値 x_1, \dots, x_n のみを取ると仮定する．集合 $\{x_1, \dots, x_n\}$ 上の関数 p_X を，

$$p_X(x) = P(X = x) \quad (1.4.1)$$

により定義する．このとき， p_X を X の**確率分布関数**（あるいは単に**確率分布**）という．また， X が確率分布 p_X に**従う**ともいい， $X \sim p_X$ のように書く．確率分布は，以下の性質を持つ（いずれも確率関数のルールと定理 1.1.1 より出る）：

- $0 \leq p_X(x_i) \leq 1$ ($i = 1, \dots, n$).
- $p_X(x_1) + \cdots + p_X(x_n) = 1$.

例 1.4.1: 確率分布の例：サイコロ投げ

再び，例 1.3.1 のように確率変数 X を定める．このとき， X の確率分布 p_X は，

$$p_X(1) = \cdots = p_X(6) = \frac{1}{6}$$

となる．

一般に,

$$p_X(x_1) = \cdots = p_X(x_n)$$

が成り立つとき p_X を一様分布という.

1.4.1 期待値, 分散, 標準偏差

期待値:

$$E[X] = p_X(x_1)x_1 + \cdots + p_X(x_n)x_n.$$

分散:

$$V[X] = E[(X - E[X])^2].$$

標準偏差:

$$\sigma[X] = \sqrt{V[X]}.$$

1.4.2 確率変数の独立性 (実現値有限個の場合)

確率変数 X, Y の同時確率分布:

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

定義 1.4.1: 確率変数の独立性 (実現値有限個の場合)

確率変数 X, Y が独立であるとは, 全ての x, y に対して

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

が成り立つことを言う.

例 1.4.2: 独立な確率変数の例: サイコロ投げ

2つのサイコロを同時に投げるとき, サイコロ1の出る目を X , サイコロ2の出る目を Y とすると, X, Y は明らかに独立である.

実際の問題では, 2つの変数に関して, 一方の変数のとる値がもう一方の変数の確率分布に影響を及ぼさないと考えられる場合, それらの変数は独立であると**仮定して**議論することが多い.

1.5 確率分布（連続実現値の場合）

この節からは、微分積分の初歩の知識を仮定する。

ここでは、確率変数 X はあらゆる実数値を取りうるものとする。また、関数

$$f_X(x) = P(X < x)$$

は微分可能であると仮定する。ここで、 f_X の導関数を p_X と書き、 X の**確率分布**または**確率密度関数**と呼ぶ：

$$p_X = f'_X.$$

このとき次が成り立つ：

$$P(X < x) = \int_{-\infty}^x p_X(t) dt.$$

確率密度関数の特徴（いずれも確率関数のルールより出る）：

$$p_X(x) \geq 0, \quad \int_{-\infty}^{\infty} p_X(x) dx = 1.$$

実現値有限個の場合と同様、 X は確率分布 p_X に**従う**とも言い、 $X \sim p_X$ と書く。

期待値：

$$E[X] = \int_{-\infty}^{\infty} x p_X(x) dx.$$

分散：

$$V[X] = E[(X - E[X])^2].$$

標準偏差：

$$\sigma[X] = \sqrt{V[X]}.$$

1.5.1 確率変数の独立性 (一般の場合)

定義 1.5.1: 確率変数の独立性

確率変数 X, Y が**独立**であるとは、全ての x, y に対して

$$P(X < x, Y < y) = P(X < x)P(Y < x)$$

が成り立つことを言う。

コメント 1.5.1

上記の独立性の定義は、確率変数の実現値が有限個の場合も、連続的な値を取る場合も含む一般の定義である。実現値有限個の場合には、確率分布関数が式 (1.4.1) の意味で定義され、上の定義 1.5.1 は定義 1.4.1 と同値になる。

1.5.2 正規分布

標準正規分布 N_0 :

$$N_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

期待値 μ , 標準偏差 σ の**正規分布** $N(\mu, \sigma)$:

$$N(\mu, \sigma)(x) = \frac{1}{\sigma} N_0\left(\frac{x - \mu}{\sigma}\right)$$

標準正規分布は、期待値 0, 標準偏差 1 の正規分布に等しい :

$$N_0 = N(0, 1).$$

定理 1.5.1: 正規分布の性質

確率変数 X が、期待値 μ , 標準偏差 σ の正規分布に従うとする :

$$X \sim N(\mu, \sigma).$$

このとき、 $P(X < x)$ は μ, σ に依らない関数 f を使って

$$P(X < x) = f\left(\frac{x - \mu}{\sigma}\right)$$

と書ける。なお、 f は単調増加関数である。

上記定理の証明は省略する。

例 1.5.1: 正規分布の例：テストの得点の分布

ある高校の2年生全体で行われた学力テストの結果は、以下の通りだったとする：

- A組：平均点 60, 標準偏差 15
- B組：平均点 70, 標準偏差 8

このとき、「90 点以上取った人数が多いクラス」はどちらと予想されるだろうか？なお、得点分布は A 組, B 組いずれも正規分布に従うと仮定する。まず、生徒の取る点数を確率変数 X で表わし、平均を μ , 標準偏差を σ としよう。仮定より X は正規分布に従うから、 $X \sim N(\mu, \sigma)$ である。ここで考えたいのは $P(X \geq 90)$ だが、これは

$$P(X \geq 90) = 1 - P(X < 90)$$

と書き直すことができる。定理 1.5.1 より、

$$P(X < 90) = f\left(\frac{90 - \mu}{\sigma}\right)$$

であるから、

$$P(X \geq 90) = 1 - f\left(\frac{90 - \mu}{\sigma}\right)$$

と書き表わすことができる。 f は μ, σ に依らない関数だったから、これで $P(X \geq 90)$ の大小関係が判別できる。 f は単調増加なので、 $(90 - \mu)/\sigma$ が小さいほど、 $P(X \geq 90)$ が大きくなることがわかる。

そこで、各クラスのテスト結果を $(90 - \mu)/\sigma$ に当てはめてみると、

- A 組：

$$\frac{90 - 60}{15} = 2$$

- B 組：

$$\frac{90 - 70}{8} = 2.5$$

となり，A 組の方が小さい．従って，90 点以上を取った人数は A 組の方が多いと考えられる．

一見すると，平均点が低い A 組の方が成績で劣っているように思われるが，この場合は標準偏差の大きさ——言い換えれば「裾野の広さ」——で平均点の低さをカバーしたのだと解釈できる．