統計・機械学習のための高校数学

統計学や機械学習によるデータ分析の理論は様々な数学に支えられて成り立っています。数学を用いて分析方法について理解を深めることで、その方法を正しく利用できたり、有用性を正確に把握できたりします。

しかしこういった数式による理論を理解するためには知識や計算スキルがそれなりに必要になるのも事実です。特に中学・高校以来数学から離れてしまった大人の方々にとってはハードルが高いものになっています。

この講座ではデータ分析を学ぼうとしている方々、学び始めたけれど数学でつまづいている方々に向けて必要な高校数学の知識を紹介します。

実は入門レベルで必要となる数学の知識は限られているので、高校 3 年分をすべて網羅するのではなく内容を絞って扱います。

1　場合の数・確率

統計学では確率の計算に基づいて様々な結論を導きます。また、条件に該当するパターンの総数を数え上げる計算は確率の計算の基礎になります。この「条件に該当するパターンの総数」を場合の数といいます。

1.1　場合の数の基礎

定理 1.1.1　和の法則：事柄 A の場合の数が a 通り、事柄 B の場合の数が b 通りあるとする。また、この 2 つは同時には起こらないとする。このとき A または B のどちらかが起こる場合の数は $a+b$ 通り。

積の法則：事柄 A の場合の数が a 通りあり、それぞれに対して事柄 B の場合の数が b 通りあるとする。このとき A の起こり方と B の起こり方のペアは ab 通り。

問題 1.1.1　(1)　さいころを 2 個投げる。目の和が 5 の倍数になるのは何通りあるか。

(2)　さいころを 2 個投げる。目の和が 4 の倍数になるのは何通りあるか。

問題 1.1.2　ある店舗で T シャツを選んでいる。色は赤、青、黄の 3 色、サイズは S、M、L、XL の 4 サイズから選ぶとする。このとき,T シャツの選び方の総数はいくつあるか。

問題 1.1.3　$(a+b)(c+d)(e+f+g)$ を展開したとき、項はいくつ現れるか。

1.2 順列と組合せ

積の法則の応用として計算できる典型的な場合の数として順列と組み合わせがあります。特に組合せの計算は二項分布など統計学の重要なポイントで何度も登場します。

定義 1.2.1 順列と組み合わせ

異なる n 個のものから r 個を選び並べる方法を順列という。この順列の総数を ${}_nP_r$ で表す。

異なる n 個のものから r 個を選ぶ方法を組合せという。この組合せの総数を ${}_nC_r$ で表す。

定理 1.2.1**（順列と組合せの計算）**

$$ {}_nP_r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!} $$

$$ {}_nC_r = \frac{n(n-1)\cdots(n-r+1)}{r(r-1)\cdots 3\cdot 2\cdot 1} = \frac{{}_nP_r}{r!} $$

問題 1.2.1 次の場合の数の計算を通して上の定理を理解せよ。

(1) 異なる 5 文字 A,B,C,D,E から 3 つを選び並べる方法の総数を求めよ。「異なる 6 文字 A,B,C,D,E,F」に増やしたらどうなるか。

(2) 異なる 5 文字 A,B,C,D,E から 3 つを選ぶ方法の総数を求めよ。「異なる 6 文字 A,B,C,D,E,F」に増やしたらどうなるか。

問題 1.2.2 次の計算をせよ。(1) ${}_8C_3$ (2) ${}_8C_5$ (3) ${}_6C_4$ (4) ${}_5C_4$ (5) ${}_5C_3$

1.3　${}_nC_r$ の性質

組合せの総数 ${}_nC_r$ について、いろいろな計算法則が知られています。これらを用いることで計算を簡単にしたり、統計学における定理を導くことができます。

定理 1.3.1　(1) ${}_nC_r = {}_nC_{n-r}$
(2) ${}_nC_r = {}_{n-1}C_{r-1} + {}_{n-1}C_r$(パスカルの三角形)
(3) $r\,{}_nC_r = n\,{}_{n-1}C_{r-1}$

あおい「うわ！いきなり難しそうな式が 3 つも出てきた！」
ほたる「数式の見かけに惑わされずに、意味と合わせて考えてみよう。」

問題 1.3.1　前節の問の結果において上の定理が成立していることを確認せよ。また、定理の証明も複数与えよ。

問題 1.3.2　A,B,C,D,E,F,G の 7 つから 4 つを選ぶことを考える。

(1) A,B,C,D から 4 つ選ぶ方法は何通りか。

(2) A,B,C,D から 3 つ、E,F,G から 1 つ選ぶ方法は何通りか。

(3) A,B,C,D から 2 つ、E,F,G から 2 つ選ぶ選ぶ方法は何通りか。

(4) A,B,C,D から 1 つ、E,F,G から 3 つ選ぶ選ぶ方法は何通りか。

(1) から (4) の結果の合計が ${}_7C_4$ に一致することを確認せよ。このことから示唆される法則について考察せよ。

1.4　二項定理

多項式の展開の計算にも組合せの考え方が現れています。

問題 1.4.1　次の多項式を展開せよ。

(1) $(a+b)^2$

(2) $(a+b)^3$

(3) $(a+b)^4$

一般に次の定理が成り立つ。

定理 1.4.1（**二項定理**）　$(a+b)^n = {}_nC_0a^n + {}_nC_1a^{n-1}b + {}_nC_2a^{n-2}b^2 + \cdots + {}_nC_nb^n$

問題 1.4.2　次の多項式を展開せよ。

(1) $(x+1)^3$

(2) $(x-1)^4$

(3) $(2x+y)^4$

1.5　確率の定義と基本性質 1

ここからは確率の内容に入ります。基本は場合の数の計算ですが、さらに発展的な内容も扱います。

定義 1.5.1　結果が偶然によって決まる実験や観測などを**試行**という。試行の結果起こることを**事象**という。

事象 A について、A が起こる確率を

$$P(A) = \frac{A\text{ が起こる場合の数}}{\text{起こりうるすべての場合の数}}$$

と定める。

例 1.5.1　さいころを 1 つ投げるとき 2 の目が出る確率は $\frac{1}{6}$, 偶数の目が出る確率は $\frac{3}{6}$

定理 1.5.1

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

A が起らないという事象を A の余事象といい $\bar{A}$ と表す。

$$P(\bar{A}) = 1 - P(A)$$

問題 1.5.1　さいころを 3 個投げるとき次の確率を求めよ。

(1)　少なくとも 1 つが 5 以上である確率。

(2)　出た目の積が偶数である確率。

1.6　確率の基本性質 2(独立な事象の確率)

定義 1.6.1　2 つの試行が互いに他の結果に影響を及ぼさないとき、これらの試行は**独立**であるという。

定理 1.6.1　2 つの試行 S, T が独立であるとするとき、「試行 S では A が起こり試行 T では B が起こる」という事象という事象が起こる確率 $P(A \cap B)$ は

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

問題 1.6.1　袋 A には白玉が 4 個、赤玉が 3 個入っている。袋 B には白玉が 2 個、赤玉が 6 個入っているとする。A から 1 個,B から 2 個の玉を取り出すとき、赤玉が 3 個とられる確率を求めよ。

問題 1.6.2　さいころを 4 回投げる。

(1) 出る目の最大値が 4 である確率を求めよ。

(2) 出る目の最小値が 3 である確率を求めよ。

1.7　条件付き確率

定義 1.7.1　事象 B が起こる条件の下で事象 A が起こる条件付き確率 $P(A \mid B)$ を

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

と定める。

問題 1.7.1　1 から 10 までの数字が 1 つずつ書かれたカードが 1 枚ずつ、計 10 枚ある。1,3,5,7,10 は赤いカード、2,4,6,8,9 は白いカードに書かれている。次の確率を求めよ。

(1) 偶数が出る確率

(2) 出たカードが赤であることが見えた。このとき偶数が出る確率。

あおい「(2) はカードが赤いってことは奇数が来そうじゃない？」
ほたる「そこが条件付き確率のポイントだね。」

問題 1.7.2 **（くじ引きは平等）**　5 本中あたり 2 本のくじがある。あおいさん、ほたるさんがこの順でひく。次の確率を求めよ。

(1) あおいさん、ほたるさんともに当たる確率。

(2) あおいさんが当たり、ほたるさんは外れる確率。

(3) あおいさんは外れ、ほたるさんが当たる確率。

(4) あおいさん、ほたるさんともに外れる確率。

問題 1.7.3 **（原因の確率）**　ある病気に罹っている確率が 0.01 であるとする。この病気を検査したい。この検査は病気に罹っている人を病気であると判定する確率、病気に罹っていない人を病気でないと判定する確率が共に 0.99 である。この検査で陽性が出たとき、実際に病気である確率を求めよ。

問題 1.7.4　箱 A には赤玉が 4 個、白玉が 1 個入っている。箱 B には赤玉 2 が個、白玉が 3 個入っている。今、どちらかの箱を無作為に選び、玉を 1 個とったところ赤玉が出た。選んだ箱が A であった確率を求めよ。

1.8　いろいろな確率分布

統計学では確率分布の概念も非常に重要です。二項分布と幾何分布という比較的扱いやすい確率分布を例に理解しましょう。

問題 1.8.1　10 本中 3 本が当たりであるくじがある。このくじを引いて確認し、元に戻す作業を 4 回くり返す。次の確率を求めよ。

(1) 当たりが 1 回出る確率

(2) 当たりが 2 回出る確率

問題 1.8.2　10 本中 3 本が当たりであるくじがある。このくじを引いて確認し、元に戻す作業を当たりが出るまでくり返す。次の確率を求めよ。

(1) 3 回で当たりが出る確率

(2) 4 回で当たりが出る確率

1.9 確率分布の期待値

期待値は平均値と呼ぶこともあります。例えば、宝くじをすべて買い占めたとした時の平均当選額は宝くじの当選金額の期待値に他なりません。確率分布の情報を調べるときに、期待値は基本的な役割を持ちます。

定義 1.9.1 **(確率変数と期待値)** さいころの出る目やくじを引いて当たりが出た本数など、確率的に値が定まる変数を確率変数といいます。

値 $X = x_i$ が得られる確率が p_i であるとする $(i = 1, 2, \cdots, n)$。このとき、X の値の期待値を

$$\sum_{i=1}^{n} x_i p_i$$

で定義する。

問題 1.9.1 1000 円が 1 本、500 円が 3 本、100 円が 6 本の計 10 本が入っているくじがある。このくじから 1 本引くとき、賞金の期待値を求めよ。

問題 1.9.2 10 本中 3 本が当たりであるくじがある。このくじを引いて確認し、元に戻す作業を 4 回くり返す。当たりの回数の期待値を求めよ。