

1.2 節 「数学基礎」 補足資料

(2024 年 8 月 30 日版)

(1) 集合とベン図

考えている何らかの対象の集まりを**集合** (set) とよぶ。集合に属する個々の対象はその集合の**元** (element) あるいは**要素**とよばれる。たとえば、元 a と元 b からなる集合は $S = \{a, b\}$ のように表される。元 x が集合 S に属することを、 $x \in S$ と書く。逆に、 x が S に属さないことは $x \notin S$ と表現する。集合に属する元の数是有限個でも無限個でもよく、後者の例としては自然数全体の集合や実数全体の集合があげられる。なお、ここでは個々の元が集合に属するかどうかだけを問題にするため、同じ元を複数含むような集合は扱わない。

集合に関する主要な概念をまとめておこう。2つの集合 S_1, S_2 があつたとき、この2つの少なくともどちらかに属する元全体の集合を**和集合** (union) とよび、 $S_1 \cup S_2$ と書く。逆に、この2つの両方に属する元だけからなる集合を**共通部分** (intersection) とよび、 $S_1 \cap S_2$ と書く¹⁾。 S_1 のすべての元が S_2 の元でもあるとき、 S_1 は S_2 に含まれるといい、 $S_1 \subseteq S_2$ と書く。このとき、 S_1 は S_2 の**部分集合** (subset) とよばれる²⁾。 S_1 のすべての元が S_2 の元でもあり、逆に S_2 のすべての元が S_1 の元でもあるとき、すなわち、 $S_1 \subseteq S_2$ かつ $S_2 \subseteq S_1$ が成り立つとき、集合 S_1 と S_2 は**等しい** (equal) といい、 $S_1 = S_2$ と書く。

また、考えている元すべてを含む集合を**全体集合** (universal set) とよぶ。ある集合 S の**補集合** (complement) は全体集合に属する元のうち、 S に属さない元の全体として定義され、 S^c や \bar{S} のように表す。なお、特殊な集合として、ど

¹⁾ 共通部分のことを積集合とよぶこともある。

²⁾ S_1 が S_2 の部分集合であることを $S_1 \subset S_2$ と書くこともあるが、この記号は \subseteq と区別して $S_1 \subseteq S_2$ かつ $S_1 \neq S_2$ となる状況 (真部分集合とよばれる) を表すために使われることもある。

んな元も属さない集合は**空集合** (empty set) とよばれ、 \emptyset と書く。

上で説明したような集合の関係は、**ベン図** (Venn diagram) を用いて表現すると直観的に理解しやすい。ベン図では1つの集合を1つの円で表し、円の交わった部分で共通部分を表す。また、ベン図が書かれている四角形全体は全体集合を表す。4つ以上の集合の関係に対するベン図の描き方も考案されているが読解が難しくなるため、3つの集合に対するベン図までがよく用いられる。

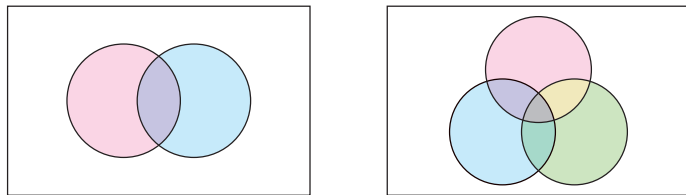


図 A.1.1 2つの集合のベン図 (左) と、3つの集合のベン図 (右)

集合の考え方は確率をより一般的に扱うために必須の概念であり、確率論の基礎としても位置付けられる。

(2) 順列, 組合せ

n 個の元からなる集合から、 m 個の異なる元を順番に選び出すことで得られる並びを**順列** (permutation) とよぶ。たとえば、集合 $\{a, b, c, d, e\}$ から選び出した (a, b, e) は順列の例である。順列では、順番を気にするので、 (a, b, e) と (b, e, a) は異なる順列として扱われる。

類似した概念として、**組合せ** (combination) がある。組合せでは、同じく n 個の元からなる集合から、 m 個の異なる元を選び出すが、元が並べられた順番は考慮しない。したがって、 $\{a, b, e\}$ と $\{b, e, a\}$ は同じものとみなされる。

順列や組合せを数え上げてみよう。たとえば、集合 $\{a, b, c\}$ から2個の元を取り出して作られる順列には、 (a, b) , (b, c) , (c, a) , (b, a) , (c, b) , (a, c) の6種類がある。同様に2個の元を取り出して作られる組合せには $\{a, b\}$, $\{b, c\}$, $\{c, a\}$ の3種類がある。もともになる集合の元の数やそこから取り出す元の数が増えると、順列や組合せを具体的に書き出すことは大変になるが、その場合の数(総数)を数え上げることは難しくない。順列の場合、もともになる集合から、まず最

初の元を選び出す必要があるが、その場合の数は n である。2 つ目の元を選び出す際には最初に選んだ元以外から選び出す必要があるため、その場合の数は $n-1$ になる。このように、選ぶことができる元の数は $n-2, n-3$ と減っていき、 m 個目を選ぶときには $n-(m-1)$ となる。そのため、 n 個の元から m 個の元を取り出した順列の場合の数 ${}_nP_m$ は、

$${}_nP_m = n \times (n-1) \times (n-2) \times \cdots \times (n-m+1)$$

と計算できる。特に、 n 個の元から n 個すべてを取り出した順列を考えると、これは n 個の元のすべての可能な並べ替えを考えることに対応しており、 n の階乗、 $n!$ が現れる。

$${}_nP_n = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1 = n!$$

階乗の記号を使うと、一般の順列の場合の数 ${}_nP_m$ は、

$${}_nP_m = \frac{n!}{(n-m)!}$$

と書くこともできる³⁾。一方、組合せの場合の数は、取り出した m 個の元が並ぶ順番を考慮せず、同じ元が含まれている組合せはすべて同一と考えるため、順列に比べて場合の数は少なくなる。 m 個の元の並ぶ順番を考えると $m!$ の場合があるので、順列の場合の数は組合せの場合の数よりも $m!$ 倍多いことになる。このことから、 n 個の元から m 個の元を取り出した組合せの場合の数 ${}_nC_m$ は

$${}_nC_m = \frac{{}_nP_m}{m!} = \frac{n!}{m!(n-m)!} = \frac{n \times (n-1) \times (n-2) \times \cdots \times (n-m+1)}{m \times (m-1) \times \cdots \times 2 \times 1}$$

のように計算される。順列や組合せの場合の数は確率統計のさまざまな場面において利用されている。

(3) 統計的仮説検定

統計的仮説検定 (statistical hypothesis test) は確率的な現象に関して得られたデータから、何らかの判断をおこなう必要がある際に用いられる便利な手法である。たとえば、あるダイエット法に体重を減少させる効果があるかどうかを判定したいとしよう。もしも、「効果があるダイエット法ならそれをおこなっ

³⁾ とくに、 $0! = 1$ であるため、 $m = n$ の場合にもこの式は正しく成り立つ。

たすべての人で大きく体重が減少し、効果がないダイエット法ならすべての人でまったく体重が減少しない」と、期待できるのであれば判定は簡単だろう。しかし、実際には、効果があるとみなせるダイエット法であってもすべての人に効果が出るとは限らず、中には逆に体重が増加してしまう人もいるかもしれない。また、このダイエット法をおこなわなくとも、自然に体重が減少する人もいるかもしれない。このため、複数名の被験者にダイエット法を実践してもらい、その前後で体重を測定したとしても、体重の増えた人と減った人が混在することがほとんどであろう。この時点で被験者の体重の平均値に減少がみられなければ、このダイエット法には体重を減少させる効果がなかったと考えてよさそうだ。では、逆に体重の平均値が減少していさえいればこのダイエット法に効果があったと考えてよいだろうか？ 実際には、そうではない。たとえば、仮に各被験者の体重がランダムに増減しているとすれば、測定前後でたまたま体重が減少していただきの可能性も十分ある。このような確率的な現象に対して、効果の有無の判定をおこないたい場合に用いられるのが(統計的) 仮説検定である。

仮説検定では一旦ダイエット法に何の効果もないことを仮定する(帰無仮説, null hypothesis)。次に、実際の実験結果で見られた以上の体重の減少がまぐれ当たりで起こる確率(P -値, P -value)を計算する。計算されたまぐれ当たりの確率が十分大きい場合には、実験で見られた体重減少は帰無仮説の下でたまたま起きたものと区別がつかない。逆に、まぐれ当たりでそんなにも体重が減少してしまう確率が非常に小さいのであれば、ダイエット法には何の効果もないのにたまたま非常に確率が低いまぐれ当たりが起きたと考えるよりも、ダイエット法には実際に効果があると考え(対立仮説, alternative hypothesis)ほうが合理的だろう。このように、仮説検定では実験結果から計算された P -値の大きさに応じて、効果がないという帰無仮説と、効果があるという対立仮説のどちらがより合理的かを判定する。このとき、どれだけまぐれ当たりの可能性が小さければ対立仮説を支持するべきかという判断基準となる閾値は有意水準(significance level)と呼ばれる。有意水準は P -値を計算する前に事前に設定しておかなければいけない。分野にもよるが、有意水準としては 0.05 や 0.01 と

いった値が使われることが多い。仮説検定の考え方は背理法に類似しており、初見では意味がわかりにくいかもしれないが、確率的なデータから判断をおこなう上では非常に有用な考え方である。ここでは紹介しないが、実際に P -値を計算する際には考えている帰無仮説に応じてさまざまな確率分布が利用される。

課題学習

- A.1.2-1** 2つの集合に対するベン図において、和集合、共通部分、また2つの集合のそれぞれに対する補集合に対応する領域を示せ。
- A.1.2-2** 集合 $\{a, b, c, d\}$ から2個の元を取り出して作られる順列と組合せを具体的に書き下し、その場合の数が ${}_4P_2$, ${}_4C_2$ とそれぞれ一致することを確認せよ。
- A.1.2-3** 仮説検定による判定をおこないたい状況を具体的に考え、まぐれ当たりの確率 (P -値) の大きさによってどのような結論になるかを議論せよ。
-