第 2 章

初等的な確率論

統計学において，集められたデータの信頼性や有効な利用を保証するのは確率論である．

そのためにまず，集合と論理について簡単なまとめをしておく．確率は一般に「事象」を扱うものであるから，事象そのものの「集まり」を概念化して扱うことは有益である．

その後，中学，高校で扱った確率について基本的な概念を掘り下げて，整理しておこう．

2.1 集合と論理代数

2.1.1 集合とは何か

あいまいさなく定義できる**要素 (element)** を集めたものを**集合 (set)** という．要素のことを元ということもある．

要素としてはどんなものでもよいので，たとえば「すべての人間の集合」は，人間の範囲をきちんと定義すれば，集合とみなせる[*1]．「すべての整数の集合」，「すべてのアミノ酸の集合」のように，要素の定義が厳密になされたものの場合には，その集合は数学的に意味がある．

一方，「美人の集合」というのは，数学的な集合としては扱えない．「美人」の定義というのは人によって，文化や時代によって様々であるから，一人の女性（たぶん美人は女性に限ると思う）がその集合に含まれるかどうかはあいまいだからである．同じように，「大きな数の集合」も数学的には意味をなさない[*2]．

[*1] 「人間の範囲をきちんと定義する」というのは実際問題としてはそう単純ではない．生きている人間のことを指すのか，そうだとすればこの瞬間に生まれようとしている人間，死ぬ間際の人間，いったいどこまでを範囲に含めるべきなのかを考え始めると厄介なことになる．そういう境界領域の問題というのは，物事を厳密に考えようとするとしばしば現われる．このように現実的な問題を数学で扱うためには，適切な区切りを付けることが必要になる．

[*2] ただし，このように境界があいまいな集合であっても，「ファジーな集合」という名の下に研究の対象となっている．

2.1.2　集合とその演算

■**集合の表現 — ベン図**

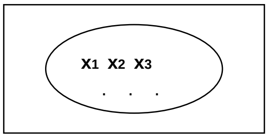
集合を A, その要素を $x_1, x_2, \ldots$ とする．これを次のように表すことにしよう．

$$A = \{x_1, x_2, \ldots\} \tag{2.1}$$

x が A の要素であることは次のように表す．

$$A \ni x,\ \text{または}\ x \in A \tag{2.2}$$

さらに，これを図にすると次のような感じになる．



ただし，以後必要がないときには，集合の要素をいちいち描かないで，単に丸や四角の「囲み」だけで集合を表すことにする．このように集合を閉じた囲みで表した図を**ベン図 (Venn diagram)** といい，以下に述べるように集合の間の関係を表現するのによく用いられる．

■**空集合**

要素をひとつも含まない集合を**空集合 (empty set/null set)** という．何も含まないものを集合というのは変な感じがするかもしれないが，数字の 0 と同じように，空集合という概念を用いることで，集合の体系的な記述が可能になるのである．空集合は，記号 $\emptyset$ で表す[*3]．

■**部分集合**

集合 A と集合 B があり，B の任意の要素が A に含まれるならば B は A の**部分集合 (subset)** であるといい，次のように表記する．

$$A \supset B,\ \text{または}\ B \subset A \tag{2.3}$$

空集合 $\emptyset$ あるいは A 自身も A の部分集合である．また，

$$A \supset B,\ \text{かつ}\ A \neq B \tag{2.4}$$

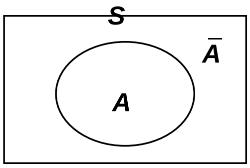
のとき，B は A の**真部分集合 (proper subset)** であるという．たとえば，犬全体の集合は哺乳類全体の集合の真部分集合になっている．

[*3] ギリシャ文字のファイの小文字 ϕ を使うこともあるが，本来は正しくない．

■全体集合, 補集合

なんらかの集合 A を想定するとき，同時に A に含まれない要素の全体 を暗黙に，あるいは明示的に想定することがしばしばある．このとき，考えられる要素すべてを含む集合を S とし，これを**全体集合**という名で呼ぶ．

S に含まれて，A に含まれない要素の集合を A の**補集合 (complementary set)** といい $\bar{A}$ で表す．$S, A, \bar{A}$ の関係をベン図で表すと下のようになる．

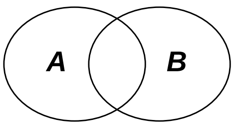


A の補集合は A に含まれない要素の集合を意味するのであるから，論理的には**否定**を意味することになる．

全体集合は，取り扱っている問題によって決まるものである．つまり，たとえば A としてすべての犬の集合をとったとき，S がすべての哺乳類の集合になるのか，すべての動物の集合になるのかは，どんな議論をしているのかによる．

■集合の二項演算

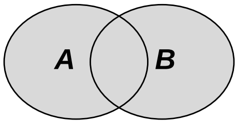
2 つの集合 A, B が与えられており，これらは共通の要素を含んでいるものとしよう．このとき，A, B の関係はベン図で下のように表される．



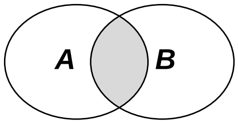
このとき，A または B の少なくとも一方に含まれている要素の集合を $A \cup B$ と表し，これを 集合 A と B の**和集合 (union)** といい，「A または B」ということもある．

また，A および B の両方に含まれている要素の集合を $A \cap B$ と表し，これを 集合 A と B の**共通部分 (intersection)**[*4]といい，「A かつ B」ということもある．下の 2 つのベン図に和集合と共通部分を示した．

[*4] 積集合ということもあるが，最近では共通部分という言い方が一般的になってきている．



$A \cup B$



$A \cap B$

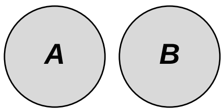
集合 A, B が共通の要素を含まない，つまり

$$A \cap B = \emptyset \tag{2.5}$$

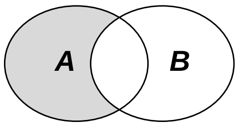
であるとき，これらは互いに**素**であるという．

集合 A, B が互いに素であるとき，$A \cup B$ を $A + B$ で表し，集合 A, B の**直和**という．

また，集合 A の要素から集合 B の要素を除いて得られる要素全部の集合を $A - B$ であらわし，集合の**差** という．



$A + B = A \cup B$

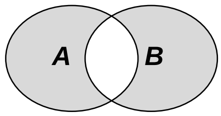


$A - B$

問題 2–1　次の関係をベン図で確かめなさい．ここでは補集合を扱っているので，A, B 両方を含む全体集合 S を考える必要がある．

$$A - B = A \cap \bar{B}$$

問題 2–2　下図のように，集合 A, B の和集合から共通部分を除いた集合を A, B の**排他的論理和 (exclusive or)** という．A, B に対して補集合，和集合，共通部分をとる演算を組み合わせることで，排他的論理和を表しなさい．



2.1.3　集合の演算規則

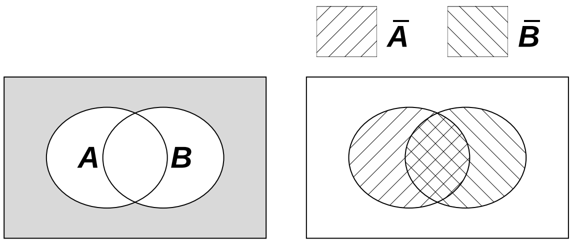
集合 A, B, C が与えられたとき，次の関係式が成り立つ．これらはベン図を描いてみれば直観的に理解できる．

$$
\begin{array}{lrcl}
\textbf{交換法則} & A \cup B & = & B \cup A \\
 & A \cap B & = & B \cap A \qquad (2.6) \\
\textbf{結合法則} & A \cup (B \cup C) & = & (A \cup B) \cup C \\
 & A \cap (B \cap C) & = & (A \cap B) \cap C \qquad (2.7) \\
\textbf{分配法則} & A \cap (B \cup C) & = & (A \cap B) \cup (A \cap C) \\
 & A \cup (B \cap C) & = & (A \cup B) \cap (A \cup C) \qquad (2.8)
\end{array}
$$

次の**ド・モルガンの法則**もよく知られている．これは論理を扱うときにも非常に重宝な関係式である．

$$
\begin{aligned}
&\overline{A \cup B} = \bar{A} \cap \bar{B} \\
&\overline{A \cap B} = \bar{A} \cup \bar{B} \qquad (2.9)
\end{aligned}
$$

下のベン図を見ながら考えれば，この式が成立することが分かる．



ド・モルガンの法則は 3 つ以上の部分集合についても成立する．すなわち，

$$
\begin{aligned}
&\overline{A \cup B \cup C \cup \cdots} = \bar{A} \cap \bar{B} \cap \bar{C} \cap \cdots \\
&\overline{A \cap B \cup C \cup \cdots} = \bar{A} \cup \bar{B} \cup \bar{C} \cup \cdots \qquad (2.10)
\end{aligned}
$$

このことは確率の計算でもよく使われるので，ベン図で表したイメージをしっかり持っておこう．

問題 2–3　50 人のクラスの中に，眼鏡をかけている人が 27 人 (集合 **A**)，携帯を持っている人が 36 人 (**B**) いる．眼鏡をかけておらず，携帯ももっていない人は 9 人だった．

1. 全体集合を S として，この状況をベン図で描きなさい．
2. 眼鏡をかけていて，かつ携帯をもっている人は何人か．

2.2　集合と確率

2.2.1　経験的確率と数学的確率

一般の数理科学の理論では，定量的な結論が得られれば，それにもとづいた現象はかなり正確に予測できる．たとえば，天体の運動のようにニュートンの運動方程式に従う現象では，次の皆既日食の日時や場所がきわめて精密に予測されている．現実が理論の意味を証言してくれるのである．

確率の場合にも，ある事象が起きる確率をきちんとした数値で与えてくれるという意味で，その結果は定量的である．しかし，その事象の起きる確率が p であったとしても，それが実際に起きるかどうかについての確実な予想は $p=0$ または $p=1$ であるとき以外にはできない．つまり確率の値が何を意味するかということと実際に起きることとの間に，正確な対応が付けられるわけではないのである．

したがって，確率という名で与えられる数値が何を意味するのかということは，議論の前提としてあらかじめ考えておく必要がある．

■経験的確率

気象観測は長年行われているので，過去のデータを多数調べることで，たとえば 8 月上旬に台風が何個上陸するかという平均値が得ることができる．もしも大規模な気候変動が起きない限りは，その平均値を使うことで，稲が被害を受ける確率を計算して災害に備えることができるだろう．プロ野球選手の打率の数字にしても，それを確率として作戦を立てることが行われている．このように多数回経験した事象から推定される確率のことを，**経験的確率**と呼ぶ．

経験的確率は，私たちには分からない未知の要因によってランダムに現われてくる現象から来るものであり，得られる数字も幅，つまり不確かさを持つことになる．しかし，そのような限界をもつ数値でもあっても，よく吟味して活用することはきわめて有用なことである．

経験的確率の考えをさらに推し進めると，もっと積極的に過去の事象を将来予測に使おうという立場も現れてくる．じゃんけんを例にとるならば，相手が出す手は 3 種類のうちのひとつだから，どの手も 1/3 の確率で出ると考えるのではなく，相手の過去の実績から「くせ」を判断して，より積極的に確率を評価していくという戦略を想定するとよい．その場合，初手に関してはどの手も 1/3 という予測を行うものの，その後は成り行きに従って使うべき確率の値は変化することになる．このような立場に拠って考えられた確率を**ベイズ確率 (Bayesian probability)** という．

ベイズ確率は，現代の IT 技術においてもスパム（迷惑）メールの判定などに活用されるなど，近年になって大きな進展を見せており，ベイズ統計の分野が新しい発展をみせている．注目される世界なので，関心のある方は情報を探ってもらうと興味深いものが見つかるだろう．

■数学的確率

理想的なサイコロやカードを使ってゲームをすることを考えると，ある事象の起きる確率は正確に計算できる．たとえば，サイコロを振って 1 の目が出る確率は 1/6 である．これは起こり得るすべての事象つまり，「1 の目がでる」，「2 の目が出る」，... ,「6 の目が出る」という 6 通りの事象のうち，ひとつの事象が実現する確率として計算されているわけである．

一般化するために，ある問題について起こりうるすべての事象の集合 S と，その部分集合をなすある特定の事象の集合 A を考えよう．S と A の要素の数をそれぞれ n, m とする． S の要素となる事象がどれも同じぐらいの確かさで起きるとするとき，A に属する事象が起きる確率 $P(A)$ は

$$P(A) = \frac{m}{n} \tag{2.11}$$

で与えられる．このように定義される確率が**数学的確率**である．

ここで，「要素となる事象が同じぐらいの確かさで起きる」という前提がなぜ成立するかということは，数学的な議論の対象にはしない．このようにある与えられた前提を出発点（公理）として，論理な操作によって体系を作り上げるのが，**公理的方法**と呼ばれる数学の基本的な構築の仕方である．

2.2.2　集合論から確率概念へ

確率を扱うときには，起こり得る事象を要素とする集合を考える．この集合のことを**全体集合**といい，また確率論では**標本空間**という言葉を用いる[*5]．2.1.2 節の部分集合を表す Venn 図で，四角の枠で表現されている集合 S は，暗黙に全体集合のイメージを持たせたものである．

確率論は集合論に基礎付けられているので，集合論の用語に対応のつく確率の用語が多い．表 2.1 にそれらの対応関係を示した．

表 2.1　集合と確率の用語の対照表

記号	集合論	確率論
S	全体集合	標本空間
	部分集合	事象
$\emptyset$	空集合	空事象
	補集合	余事象
	和集合	和事象
	共通部分	積事象
	互いに素	排反

[*5] 数学的に厳密な確率論は旧ソ連のコルモゴルフが公理論的に定式化したものであるが，ここでは数学的な厳密性を避けて，直観的にわかりやすい説明を取っている．数学としての確率論に興味のある人は，より専門的な入門書をあたっていただきたい．

2.2.3　基本的な確率の公式

事象を A あるいは $A_1, A_2, \ldots$ で表す．起こり得るすべての事象の集合を標本空間と呼び，S で表すことにしよう．このとき，確率 $P(A)$ は次の定義に従うものとして定める．

■確率の基本定義

標本空間 S は考えられるすべての事象の集合なので，その実現確率は 1 である．

$$P(S) = 1 \tag{2.12}$$

$A_1, A_2, \ldots$ が互いに排反であるとき，確率は足し合わせられる．

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \cdots \tag{2.13}$$

このとき直和の記号を使えば，次のように簡単に書ける．

$$P(A_1 + A_2 + \ldots) = P(A_1) + P(A_2) + \cdots \tag{2.14}$$

ここからいくつかの関係が導かれる．

■空事象の確率

$$P(\emptyset) = 0 \tag{2.15}$$

これは直観的にも明らかだが，背理法で考えれば簡単に証明できる[*6]．

■余事象の確率

次の関係はとても重要で，しょっちゅう使われる．これは $S = A + \bar{A}$ からすぐにわかる．

$$P(\bar{A}) = 1 - P(A) \tag{2.16}$$

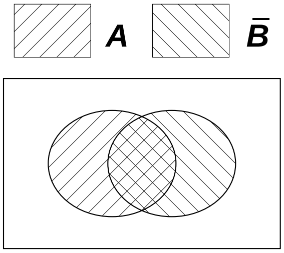
■和事象の確率

事象 A と B が与えられたとき，次の式が成り立つ．

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.17}$$

下のベン図で考えてみよう．

[*6] もし $P(\emptyset) > 0$ であったとしよう．S と $\emptyset$ は共通の要素を持たないので互いに排反である．したがって，式 (2.13) より，$P(S \cup \emptyset) = P(S) + P(\emptyset) > 1$ となり，また $S \cup \emptyset = S$ であるから，$P(S) > 1$ となって，式 (2.12) に反する．



$P(A)+P(B)$ は共通部分 $A \cap B$ を 2 回数えてしまっているので，その部分を引いているわけだ．

2.2.4　条件付き確率

ある事象 A が起きているとした上で他の事象 B が起きる確率を $P(B|A)$ で表し，このような確率を**条件付き確率 (conditional probability)** という．$P(B|A)$ については式 (2.18) が成り立つ．図 2.1 を使ってこのことを説明しよう．

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{2.18}$$

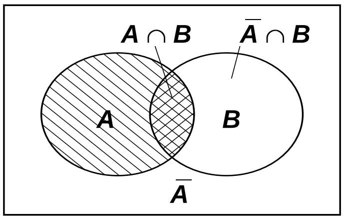


図 2.1　条件付確率の意味を Venn 図で表す

図のそれぞれの集合に含まれている根元事象の数が次のようになっているものとしよう．すなわち，全体集合に n，A に n_1，B に m，$A \cap B$ に m_1，$\bar{A} \cap B$ に m_2 のように．このとき式 (2.18) の左辺は，A に含まれる事象のうちで B に含まれている事象の起こる確率だから，次のようになる．

$$P(B|A) = \frac{m_1}{n_1}$$

右辺の分母と分子はそれぞれ，すべての事象の中で A に含まれる事象の起こる確率，すべての事象の中で $A \cap B$ に含まれる事象の起こる確率を使って，次のように表すことができる．

$$P(A) = \frac{n_1}{n}$$

$$P(A \cap B) = \frac{m_1}{n}$$

したがって，式 (2.18) が成り立つことになる．ただしこの式は，図を見て考えれば直観的に理解できるものである．

2.2.5　乗法定理と独立事象の公式

式 (2.18) を次のように変形してみよう．

$$P(A \cap B) = P(A)P(B|A) \tag{2.19}$$

この式は，2 つの事象がいずれも起こる確率を求めるための公式になっており，確率の**乗法定理**と呼ばれる．

ここでもしも次の式が成立するとしよう．

$$P(B) = P(B|A) \tag{2.20}$$

すると，式 (2.19) は次のように書き換えられる．

$$P(A \cap B) = P(A)P(B) \tag{2.21}$$

この意味を考えてみよう．$P(B) = P(B|A)$ が成り立つということは，**事象 A がすでに起きていたときに事象 B が起きる確率 $P(B|A)$ が，A が起きたか否かを問わず B が起きる確率 $P(B)$ と等しい** ということである[*7]．たとえば，2 組のカードのセットがあって，一方から引いたカードがハートであった (事象 A) として，その後で別のセットから引いたカードがハートではない (事象 B) 確率はどうなるだろうか？常識的に考えて，2 組のカードの山同士は無関係なのだからその確率は変わるはずがない．

2 つの事象 A と B があって，A が起きたことが B の起きる確率に影響を与えず，その逆も成り立つものとする．このとき，事象 A と B は**独立**であるという．2 つの独立事象が両方とも起きる確率は，それぞれの起きる確率の積に等しい．

ただし数学的には，式 (2.21) が成立するときに，事象 A, B が独立であるという[*8]．

2.2.6　独立性，因果関係，オッズ比

前節の話を，こんな例を用いて考えよう．

> ある日，犬を見た人が，たまたま風邪を引いたとする．となると，人によっては「自分が風邪を引いたのは，犬を見たことと関係があるのではないか？」という疑問にとらわれる人も出てくるだろう．それでは，この人にとって犬を見るという事象 (A) と風邪を引くという事象 (B) とは関係があるのかどうか，どうやって判断すればよいのだろうか．

そのためには十分に多数回の経験をしてみて，その上で確率を考えて見ることになる．

[*7] 論理的には，式 (2.20) だけではなく A と B をひっくり返した形の $P(A) = P(A|B)$ も成り立たないと，A, B が独立であることはいえない．しかし，片方が成り立てばもう一方も同時に成り立つことを，簡単な計算で確かめることができる．

[*8] 確率を現実の現象を扱う道具として使う立場からは，2 つの事象が独立であるかどうかは，それらを引き起こす原因の間に何らかの関係があるかどうかという観点で議論される．たとえば，ある場所で雪が積もることとそこで人が道で転ぶことの間には間違いなく関係がある．しかし，数学はそのような個々の事情から離れて抽象化された議論をするものであるから，「事象 A と B が独立である」という命題を式 (2.21) が成立することの必要十分とみなすのである．

さて，仮に犬を見ると風邪をひきやすいということがあったとしたら[*9]，次のような関係が成立することになるだろう．左辺は犬を見たとして，風邪をひくという条件付き確率で，右辺は犬を見たかどうかとは無関係に風邪をひく確率だ．

$$P(B|A) > P(B) \tag{2.22}$$

これに式 (2.18) の条件付き確率の定義の式を代入すると，

$$\frac{P(A \cap B)}{P(A)} > P(B)$$

より，

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

となる．つまり，事象 A と B がどちらも起きる確率は，それぞれが起きる確率の積よりも大きいということになる．まさにこれは，2 つの事象の間に何らかの因果関係があることを感じさせる．

一方，式 (2.22) で等号が成立していたらどうだろう．それが意味することは，**犬を見たときに風邪をひく確率と，犬がいようがいまいが風邪をひく確率は変わらない**ということだ．つまり，犬を見るという事象と風邪をひくという事象は独立であるということになる．そしてこのときには，

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

が成立する．

つまり，**2 つの事象の実現が独立であるときにのみ，事象が両方とも実現する確率がそれぞれの事象が起きる確率の積となる．**

■オッズ比と因果関係

次に，何かの病気にかかった人にある薬を飲んでもらって，その効果を検証するような場合を想定しよう．このとき，被験者は全員が「薬」を飲まされるが，その中には偽薬があって，本物を飲む人は一部である．本物を飲む事象を A とし，また B の方は，本物であれ偽物であれ，薬を飲んで一定の日数後に治癒したという事象であるとする．

十分に多数の被験者を使って得た，この試験の結果が図 2.2 のようになったとしよう．与えられている a, b, c, d を使うと，次のように確率が得られたことになる．

[*9] もちろんそんなことはないだろうが，縁起をかつぐ人はそんなふうに思い込むものだ．

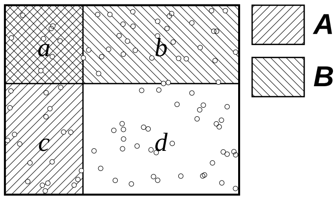


図 2.2　オッズ比と条件付き確率を考えるための図．矩形の領域は，左上から時計回りに $A\cap B$, $\overline{A}\cap B$, $\overline{A}\cap\overline{B}$, $A\cap\overline{B}$ の集合を表し，それぞれに含まれる根元事象の数は a, b, c, d である．

$$\begin{aligned} P(A) &= \frac{a+c}{a+b+c+d} \qquad (2.23)\\ P(B) &= \frac{a+b}{a+b+c+d}\\ P(A|B) &= \frac{a}{a+b}\\ P(B|A) &= \frac{a}{a+c} \end{aligned}$$

もしも，このときに事象 A と事象 B が独立だったとすると

$$P(B|A) = P(B)$$

が成立するので，式 (2.23) 以下を使って整理すると次の式が得られる．

$$\frac{a}{c} = \frac{b}{d} \qquad (2.24)$$

この式の意味を考えよう．$\frac{a}{c}$ という値は，本物の薬を飲んでいたとして，治癒した数を a としなかった数 c で割ったものである．この値を**オッズ** (odds) という．もう一方の薬を飲まなかった方についても．同じようにオッズ $\frac{b}{d}$ が定義される．

薬を飲んだ場合のオッズと，飲まなかった場合のオッズとの比の値 (前の数を後の数で割った値) を**オッズ比** (odds ratio) という．つまり，

$$\text{オッズ比} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$$

ここで扱った場合では薬の効果がなかったことを想定していて，そのときオッズ比は 1 になっている．もしも薬の効果があったとすれば，飲んだ方のオッズは大きくなり，飲まなかったほうは変わらないはずだから，オッズ比は 1 よりも大きくなることになる．

このように，オッズ比が 1 よりも大きいかどうかは，治療薬の客観的な効き目を判定する上で重要な指標になる．

例題 2–1　ある疾患の新しい治療薬の効果を試すために，100 人の罹患者に対して処方したところ，20 人が治癒した．そのことをもってこの薬には効果がないと判断してよいか．

一見すると，この薬では治ったのは全体の 1/5 だから，効果がなかったようにも感じられる．ちなみにこの場合のオッズは $20/80 = 0.25$ である．しかし，薬を飲ませなかった場合にどの程度の人が治癒するかについてのデータがないので，これだけで効果の有無を判断することはできない．客観的に効果を判定するためには，本物を投薬された人と偽薬を飲まされた人の両方でデータをとってみて，オッズを比較することが必要だ．

その結果が右の分割表のようになったとする．

	投薬	偽薬
治癒	20	11
非治癒	80	87

オッズ比を計算してみると，次のように 1 よりもかなり大きい値になるので，この薬は有効であると推測することができる．

$$\frac{20}{80} \div \frac{11}{87} = 1.98$$

実際の臨床試験においては，被験者も実験者も本物の薬と偽薬のどちらを扱っているか分からないように実施する．これを**二重盲検法** (double blind test) という．

さらに，試験データは罹患者全体を母集団としてそこから抽出して得たものと考えられるので，結果は統計的ゆらぎをもつことになる．そこで真に有効かどうかを判定するための統計的な検定を行う必要がある．それについては第 8 章で詳しく学ぶ．

2.2.7　ベイズの定理

条件付確率については，**ベイズの定理 (Bayez's theorem)** という実用的にも重要な定理が知られている[*10]．

簡単なケースについて考えてみよう．事象 A, B が互いに排反で，かつ標本空間を尽くしているとする．すなわち，次のように S が A, B の直和になっているとする．

$$S = A + B$$

このとき，A, B それぞれの下に，ある事象 E が起きる確率，

$$P(E|A), P(E|B)$$

が知られているとする．

*10　「なんとかさんの定理」というと，たいていは著名な数学者の名前から来ているものだが，ベイズという人は 18 世紀初めにイギリスに生まれた牧師で，数学者としては無名だったようだ．彼は条件付き確率の問題の特定の場合について考察を遺しており，死後にそれは発表された．

今，E が起きたとして，それが A によるものである確率 $P(A|E)$ は次のようになる．これがベイズの定理である。この定理は，事象 $A, B, C, \ldots$ と拡張してもそのまま成立する．

$$P(A|E) = \frac{P(A)P(E|A)}{P(A)P(E|A) + P(B)P(E|B)} \tag{2.25}$$

ここで想定している状況について例を挙げて説明しておこう．飲酒の習慣があるか，そうでないかという前提条件の違いは，適当な区切りを設定すれば，お互いに排反な事象とみることができる．そこでこれらをそれぞれ A, B としよう．飲酒の習慣は，当然のこととして健康への影響があると思われるので，ある病気にかかる事象を E とすると，お酒を飲む人がこの病気にかかる確率 $P(E|A)$ と飲まない人がかかる確率 $P(E|B)$ には違いがあるはずだ．たいていの病気ではたぶん $P(E|A) > P(E|B)$ つまり，お酒は体に悪いということになるのかも知れない．

さて，**もしもある人を任意に選んで，その人がこの病気にかかっていたとしよう．この人がお酒のみである確率はどれだけか？**この確率は，条件確率 $P(A|E)$ で与えられる．これがベイズの定理で設定している状況である．

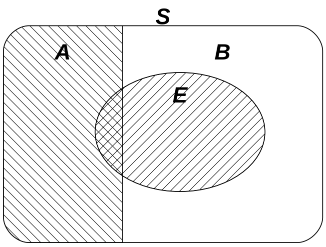


図 2.3　ベイズの定理の説明図

図 2.3 のベン図を使って考えてみよう．状況を見やすくするために，条件付き確率の説明に登場した式 (2.18) の $\bar{A}$ を B に，B を E に置き換えて少し変形すると，次の最初の式が得られ，2 つめの式も同様に得られる．

$$\begin{aligned} P(A \cap E) &= P(A)P(E|A) \quad (A \text{ と } E \text{ の共通部分が実現する確率}) \\ P(B \cap E) &= P(B)P(E|B) \quad (B \text{ と } E \text{ の共通部分が実現する確率}) \end{aligned}$$

図 2.3 の楕円形で区切りの左側と右側が，それぞれ式 (2.26) の $P(A \cap E)$, $P(B \cap E)$ に相当している．この 2 つを合わせたうちに占める $P(A \cap E)$ の割合が，E が実現しているときにそれが A によるものである確率 $P(E|A)$ であるから，式 (2.25) が得られることになる．

なお，式 (2.25) のベイズの定理は，簡単な計算で次のように変形することができる．

$$P(A|E) = \frac{P(A)P(E|A)}{P(E)} \tag{2.26}$$

この式のほうが見かけ上簡単であるために，本によってはこちらを紹介しているものもある．しかし，確率 $P(E)$ の値は $P(E|A)$ や $P(E|B)$ のように現実のデータとしては想定しにくいので，必ずしも使いやすいとは言えない．

2.2.8 ベイズの定理の例

A, B の二つの農場でトマトを市場に出荷していて，消費者はそのどちらかだけを購入しているものとする．A 農場では毎日 800 個のトマトを出荷し，そのうちの 5% は虫食いである．B 農場では毎日 2000 個のトマトを生産していて，そのうちの 12% が虫食いであるとする．

今，トマト 1 個を無作為にとって，それが虫食いだったとする．このトマトが A 農場のものである確率はどうなるか．

A 農場のトマトである確率を $P(A)$, A 農場のトマトであって虫食いである確率を $P(E|A)$, 虫食いのトマトが A 農場のものである確率を $P(A|E)$ とする． B 農場のほうについても同様に定義する．すると，ベイズの定理から次のように結果が得られる[*11]．

$$\begin{aligned} P(A|E) &= \frac{P(A)P(E|A)}{P(A)P(E|A)+P(B)P(E|B)} \\ &= \frac{\frac{800}{800+2000}\times 0.05}{\frac{800}{800+2000}\times 0.05+\frac{2000}{800+2000}\times 0.12} \\ &= 0.143 \end{aligned}$$

2.2.9 ベイズの式を使わない考え方

ベイズの公式 (式 2.25) は抽象的であってやや理解しづらいし，また逆に式に数値を当てはめるだけで中身も分らないで操作するということにもなりやすい．もっと具体的な考えに立って上の問題を解いてみよう．

問題のような状況で，トマトを 100 個，無作為に買ってきたとしよう[*12]．そのうち，何個ずつが 2 つの農場から出荷されたものかを考えると，その期待値は次のようになる．

[*11] この種の計算を行う時には，$\frac{800}{2800}$ のような分数は計算しないで進めること．そうすれば約分できて計算が劇的に簡単になる！

[*12] 100 のかわりに他の考えやすい数でもかまわない．

$$\begin{aligned} \text{A 農場産の個数} &= 100 \times \frac{800}{800+2000} = 28.571 \\ \text{B 農場産の個数} &= 100 \times \frac{2000}{800+2000} = 71.429 \end{aligned} \tag{2.27}$$

これらのうちの虫食いの数はそれぞれ，

$$\begin{aligned} &\text{A 農場産のうちの虫食いの個数} = 28.571 \times 0.05 = 1.4286 \\ &\text{B 農場産のうちの虫食いの個数} = 71.429 \times 0.12 = 8.5715 \end{aligned} \tag{2.28}$$

したがって，虫食いのうちで A 農場産のものである確率は次のようになる．

$$\frac{1.4285}{1.4285+8.5715} = 0.143 \tag{2.29}$$

このように，ベイズの公式に当てはめたのと同じ結果が得られた[*13]．これならベイズの定理を覚えていなくても解ける．

問題 2–4　疾病検査の信頼性: ある病気に感染しているかどうかを調べる検査法がある．この検査によると，被験者が非感染者なのに陽性になる確率が 1.5%，感染者なのに陰性になる確率が 0.5% ある．この病気は人口の 2% が感染しているものとしよう．ある人が陽性と診断された．この人が本当に感染者である確率はどれだけか？

[*13] これは当然のことで，このやり方を一般化したのがベイズの定理である．

2.3　認識と確率

数学の分野の中では，確率はもっともよく話題にされる．しかしながら，確率に関する人間の直感にはバイアスがかかっていて，しばしば誤ったイメージを持ってしまう．数学的な内容からは逸脱するが，確率をめぐる人間の認識の問題について考えておこう．

2.3.1　一様な確率とは何か

■野球解説者の確率論

「この打者の打率は 2 割 5 分で，これまで 3 打席三振だったから，次はヒットを打つはずですね」という野球解説者はよくいるものだ．それは正しいのだろうか．打者がコンスタントに打率を維持している場合，ヒットとアウトはどのように現われるのだろう．

この疑問を確かめるために，確率 1/4 の割合で H が，残りは O が合計 50 回出現する簡単なシミュレーションを行ってみた．その結果は次の通りである．

```
O H H O O H O O O O O O O H H H O O H H O O H O O O H O H O O O H H H
O O H O O O O O O H O H O O O O H O O H O O O O O O O O O H O O O O O
O O O O H O O O O O O O O O O O O H H O O O O O O H O O O O O O O O H
```

これを見てどう思うだろうか．きっと，「案外ヒットがかたまっているものだなあ」と感じるのではないだろうか．逆に後半になると今度はなかなかヒットが出ないスランプ状態も出現している．このように，確率そのものは一定であっても，実際に出現する事象のほうはかなり偏りを見せることになる．人はこれらを，「つき」，「スランプ」，「運に見放された状態」などと呼ぶことがあるわけだ．しかしこのシミュレーションに見るように，これらは確率の自然な表れなのである．

問題 2–5　0 から 9 までの数字を，なるべくでたらめだと思うやりかたで 50 個書き並べなさい．その後，連続して同じ数が出現する確率を計算してみて，自分の「くせ」を検証してみなさい．

今度は別の実験結果を示そう．図 2.4 は 1/10 の確率でマス目に黒石を置き，残りは白石を置いたみたところをシミュレーションでやってみたものである．これを見ると，黒い石の間にまるで引っ張りあう力が働いているかのような「意味ありげなかたまり」がたくさん現われている．

■ランダムさはなぜ意味ありげに振舞うのか

以上のように，ある確率のもとにランダムに起きるはずの現象は，人間にとってはむしろある種のパターンを感じさせることが多い．この意外さはどこからくるのだろうか．

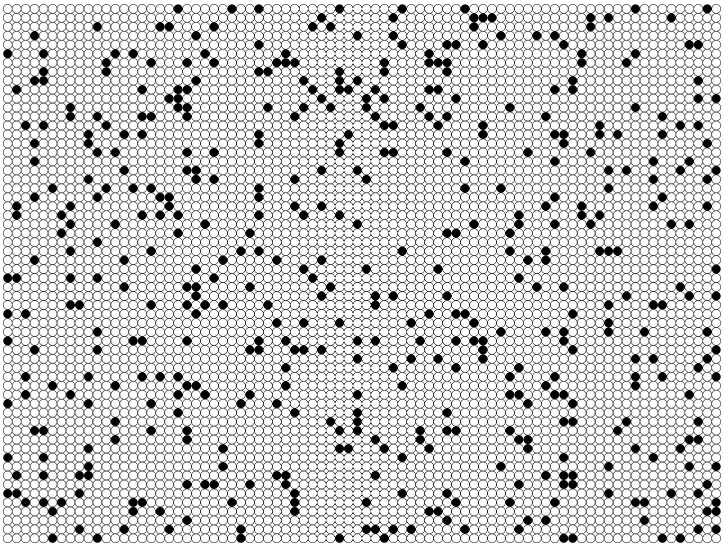


図 2.4　白 9 個に対して黒 1 個の割合で混ざった碁石をランダムに敷き詰めてできたパターン

打率が 0.25 の打者の例を考えてみよう．いま，この打者が 3 連続でアウトになったとする．さて次に起きることとして，次の 2 つの予想のどちらが正しいのだろうか．

確率の「流れ」を考える人　打率 0.25 ということは，4 打席に 1 回はヒットになるということだ．今 3 打席アウトになったところだから，そろそろヒットになってもいいはずだ．

確率は変らないと考える人　打率 0.25 ということは，どんな状況でもその次がヒットになる確率が 0.25 だということだ．だから次もアウトの確率の方が高い．

実際にシミュレーションの結果をながめてみて，どちらの予想の方が的中したのかを調べてみよう．アウトが 3 回連続している場合に，その次がアウトになっている場合とヒットになっている場合を数えてみるのである．調べてみると，アウト 3 回連続の後でアウトになっている回数は 23 回，ヒットになっている回数は 10 回となっていて，どちらかというと後の予測の方が的中している．以上を次のようにまとめておこう．

●一様な確率では，次の事象はその前の影響を受けるわけではない．

●確率的に起きる現象は，「つき」，「はずれ」とか意味ありげなまとまり方を見せることが多いが，それは確率的な現象として自然なのである．

●人間は，一様な確率を，規則的な間隔で現象が起きることと勘違いしやすい，

2.3.2　降水確率 70% の意味は

朝のニュースを見ていたら，京都市の午後の降水確率が 70% であるという予報が流された．これが意味することは，次のうちのどれにもっとも近いのだろうか．

1. この日の午後には京都市の約 70% の地域に雨が降る．
2. この日の午後に京都市のどこかで雨が降る確率は 70% である．
3. この日の午後，京都市のどこかにいたとすると，雨に遭う確率は 70% である．

それぞれについて考えてみよう．1. が正しいとしたら，京都市民のうち 70% が雨に降られる（市内の人口密度が均一だったとして）ことになるからなんとなく当たっているような気がするかもしれない．しかし，天気予報では，地域の中で雨が降る場所の面積比率を予測しているわけではない．そもそもそんな観測自体できるはずがない．現在の気象状況と過去のデータから，その市の観測地点で雨が降るであろう確率を計算しているのである．

2. は明らかにおかしい．「京都市のどこかで雨が降る」事象の余事象は，「京都市のどこでも雨が降らない」ということである．これは市内のどこかで 1 滴でも雨が降ったら成り立たなくなるわけだから，少しでも雨模様が予想される日にはその確率はほとんどゼロのはずだ．つまり，ほんの少しでも雨が降りそうな日には，どこかで雨が降る確率はほぼ 1 になってしまう．降水確率が 0 になるのは，雲ひとつない快晴のときだけということになり，70% などという中途半端な確率はほとんど現れなくなるだろう．

3. について考えてみよう．降水確率 70% という予報は，過去の観測データと現在の気象状況，それにシミュレーションによる予測とから，観測地点（気象台や測候所）で降水がある確率が 70% としているのである．大雑把に言えば，ある打者が出たらヒットになる確率がどうなるかを，その打者の打率で考えるのと似たようなものである．降水確率が 70% という予想が出た日を多数回調べれば，そのうちの 70% で予報が的中しているという，経験的確率を，天気予報として発表しているといってもよい．

3. で示された解釈は，観測地点ではなくて市内の任意の場所を考えているので，多少の違いはあるが，予報というものの性格上，観測所と地理的に近い範囲ではほぼ同じ予測を適用できるとしているのだ．だからこの解釈は妥当である．

2.3.3　ホームランボールはだれかに当たる

宝くじで一等をとった人がいると，その人には特別な神通力でもあるかのように思われることがあるし，一等が出た宝くじ売り場には「ここで一等賞が出ました！」という宣伝が貼り出されて，いかにもその売り場が特別であるかのように装うものである．

昔の知り合いに全く予想もつかない偶然で再会したりすると，「不思議な力」を感じたり，「赤い糸」を想像したりするのもよくある話だ．何回かの行事の外出のたびに雨に降られると，「雨男」，「雨女」にされて，その人は雨を呼ぶ力を持っていることになるというのもよくある．これはいわゆる**ジンクス**のたぐいである．

このように低い確率の事象が起きると，何かしら驚きや必然のようなものを感じるのが人間の感覚である．このことについて考えておこう．

■千人にひとりが「必ず」当たる

大勢で勝ち抜きじゃんけんをして，10 回勝ったら景品がもらえるイベントがあったとしよう．あなたは自分がその景品をもらえることを期待するだろうか．確率は 1/1024 であるから，たいていの人は期待しないだろう．しかし，それでも約千人にひとりの「だれか」には必ず当たるのである．各人にとっては期待できない偶然であり，一方全体ではかならず誰かに当たる必然として，景品が当たるという事象がおきるわけだ．

さてこのとき，たまたま当たった人はなんらかの意味で「特別な人」なのだろうか？私たちはそのように思うことがしばしばあるし，そのような人を「運のよい人」とも呼ぶ．

しかし，もともと「運のいい人」がいるわけではない．誰が運のいい人になるのかは，決して事前には予測できないし，最初から決まっているわけでもない．あくまで結果論としてだけ，「運がよかった人」は存在するわけで，その人は確実に現われる（ただしまずあなたではないだろう）．当たった人も，外れたあなたも，事前の状態においてはまったく平等なのである．

> 運がいいとか悪いとか，人はときどき口にするけど，そういうことって確かにあると，あなたを見ててそう思う（さだまさし「無縁坂」より）

「運がいい人」をうらやむ人はいるものだが，たまたま当たった人であっても，自分がまさかそういう立場になれるとは思いもしなかったはずだ．低い確率の幸運など当てにしないで，つまり自分が運のいい人になれるだろうとは期待しないで，自分の行動を組み立てることしか，私たちにできることはない．もちろん，たまたま幸運に恵まれることはたいていの人の人生の中で何回か起きるわけだから，それを活かすか無駄にするかは，努力や心構えの問題だ．

2.3.4　偶然はだれにもコントロールできない

念力とかサイコキネシスと呼ばれるオカルト系の話の小道具がある．たとえば目力（めぢから）で机の上の鉛筆を転がしたり，何ら力を加えないのにスプーンが曲がったりするあれである．これらは全部インチキであるから，信用してはいけない．

それでも幸運を授けてくれる護符とか，あるいは強く信じれば奇跡が起きるという信念

は，この社会に根強いものがある．前世の自分や背後霊が自分の現在に何らかの影響を与えて，それらに祈って働きかけることで自分の運を変えることができるという言説も根強い．それは妥当だろうか？たとえば加持祈祷でガンは治るだろうか．

ガンの治療というのは，手術や放射線などでガン組織を取り除いて，他の病巣が残っていないか，あったとしても制圧可能な程度であることを期待するものだ．これはいわば賭けであり，治る，治らないという 2 つの排反事象の割合を少しでもよい方に動かすべく治療を行うのである．実際にどちらの事象が発現するかは，知りようがない．このような状況に直面して，加持祈祷や護符のようなものを求める人は多い．

そのような行為が不安を鎮めるために役立つことはあるだろう．また死の不安に対してどのような行動をとるかを批判することは，慎みのない行為である．しかしそれでも，人間の情念がいかに強くとも，それだけで事象の発現を左右することはできないのである．ガンを避けるためにできることは，ガンを招く生活要因を避けること，早期発見に努めること，万一ガンが見つかったら最善の治療を選択すること，— これらができることのすべてである．ガンの発生確率や再発の確率を引き下げることができる合理的な努力をするしかない．

それでも不運に見舞われるかもしれないというところが，不安の理由なのだが，それについては，「悪運を受け入れざるを得ないかもしれない」という覚悟が必要なのだ．つまり先人の言葉を借りれば，人事を尽くして天命を待つしかないのである．

以上のように，確率の問題を考えることは，リスクの中に生きる存在である私たちの行動の仕方を問うことでもある．簡単にまとめておこう．

- ●私たちにできるのは，確率全体を増減するように状況を変えるための合理的な働きかけや努力だけだ．
- ●その都度どちらの事象が偶然に現われるのかをコントロールする方法はない．
- ●超自然的に偶然を操る存在（神とか霊とか）もない[*14]．

【章末問題】

問題 2–6　式 (2.17) を使って，52 枚のカードから 1 枚を引いたときに，それが黒であるか偶数であるか，少なくともどちらかである確率を求めなさい．ただし，クイーンは偶数

[*14] ここで言っているのは，偶然を操って現実の世界の現象を左右するような存在としての神や霊はないということである．ピッチャーの球のコースは予測不可能だし，打者の身体コントロールがたまたまそれと合致したときにヒットが生まれるわけだから，それには毎回の偶然が大きく左右する．そこでサイコロをちょいと動かしてくれる神やら霊などはないのだ．私たちにできるのは，普段の訓練とその場の集中によって打球がヒットになる事象の集合を広げることだけである．まして祈ったら現世の利益が得られるなどということはない．ただし，人が超自然的な存在に対する観念や信仰をもつことをここで否定しているわけではない．信じることから得られる心の安定や健康への影響を大事だと思う人はいるかも知れない．

には含めないものとする．

問題 2–7　ある学校のクラスの学生で，誕生日が一致する人がいる確率を考える．何人以上のクラスで，この確率が $1/2$ を超えるか．