確率・統計講義ノート

緑川章一

さあ進もう。神々の示現と確率・統計が呼んでいる世界へ。
賽は投げられた。

目次

第1章　確率の基礎

1.1　確率の公理

　標本空間を Ω とし、その部分集合を要素とする集合を $\mathcal{F}$ とする。$\mathcal{F}$ の要素 A に対して実数関数 $P(A)$ を定義し、この値 $P(A)$ を事象 A の確率と言う。それは次の性質を満たす。

1. $0 \leqq P(A) \leqq 1$
2. $P(\Omega) = 1$
3. 互いに共通部分を持たない部分集合 $A_1,\ A_2, \cdots,\ A_n$ に対して、
   $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$
   が成り立つ。

1.2　条件付確率

一般に、標本空間 Ω の部分集合 $A,\ B$ が与えられていて、かつ $P(B) \neq 0$ の場合に、

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

を事象 B が起ったときの事象 A の**条件付確率**と言う。言い換えると、条件付確率 $P(A|B)$ とは B が起ったと分った時に、それが A である確率である。

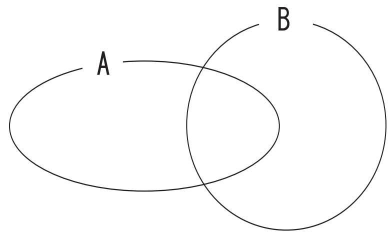

図 1.1: 共通部分と条件付確率：事象 B のもとで事象 A の起こる確率は、$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ で与えられる。

1.3　ベイズの定理

$\{A_1,\ A_2,\ \cdots,\ A_n\}$ を Ω の分割とする。すなわち、

$$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega, \quad \text{かつ} \quad i \neq j \text{ のとき } A_i \cap A_j = \emptyset$$

である。このとき、

$$P(B|A_i) = \frac{P(B \cap A_i)}{P(A_i)}$$

ゆえに、

$$P(B|A_i)P(A_i) = P(B \cap A_i)$$

一方、

$$B = B \cap \Omega = \sum_{k=1}^{n} B \cap A_k$$

だから、

$$P(B) = \sum_{k=1}^{n} P(B \cap A_k) = \sum_{k=1}^{n} P(B|A_k)P(A_k) \tag{1.1}$$

が成り立つ。

　また、

$$P(A_i|B) = \frac{P(B \cap A_i)}{P(B)}$$

である。ここで、右辺の $P(B)$ 分母に (1.1) 式を代入すると、

$$P(A_i|B) = \frac{P(B \cap A_i)}{\displaystyle\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

を得る。

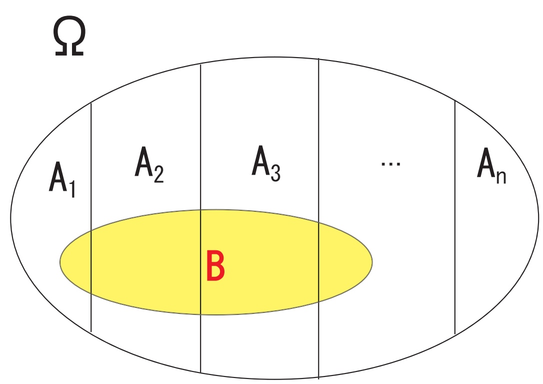

図 1.2: ベイズの定理

ベイズの定理

事象 B は、$P(B) > 0$ を満たすものとする。$\{A_1,\ A_2, \cdots, A_n\}$ を標本空間 Ω の分割とする。すなわち、$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$ で、任意の $i \neq j$ について $A_i \cap A_j = \emptyset$。このとき、

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\displaystyle\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

が成り立つ。$P(A)$ を事前確率、$P(A_i|B)$ を事後確率と言う。

ベイズの定理の応用について考えよう。標本空間 Ω が 2 つの集合 $\{A_1,\ A_2\}$ に分割されているとする。Ω は、また、 2 つの集合 $\{B1, B2\}$ を含むとすると、

$$
\begin{aligned}
B_1 &= (B_1 \cap A_1) \cup (B_1 \cap A_2) \\
B_2 &= (B_2 \cap A_1) \cup (B_2 \cap A_2)
\end{aligned}
$$

ところで、条件付確率の定義から

$$
\begin{aligned}
P(B_1|A_1) &= \frac{P(A_1 \cap B_1)}{P(A_1)} \\
P(B_1|A_2) &= \frac{P(A_2 \cap B_l)}{P(A_2)}
\end{aligned}
$$

が導かれるが、これらを用いると、

$$
P(B_1) = P((B_1 \cap A_1) + (B_1 \cap A_2) = P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2)
$$

となる。ところで、$P(A_1|B_1) = P(A_1 \cap B_1)/P(B_1)$ だから、

$$
P(A_1|B_1) = \frac{P(B_1|A_1)P(A_1)}{P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2)}
$$

とも書ける。同様にして、

$$
\begin{aligned}
P(A_2|B_1) &= \frac{P(B_1|A_2)P(A_2)}{P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2)} \\
P(A_1|B_2) &= \frac{P(B_2|A_1)P(A_2)}{P(B_2|A_1)P(A_1) + P(B_2|A_2)P(A_2)} \\
P(A_2|B_2) &= \frac{P(B_2|A_2)P(A_2)}{P(B_2|A_1)P(A_1) + P(B_2|A_2)P(A_2)}
\end{aligned}
$$

これは、図 (1.3) 見ると分かりやすい。このような書き方は、クロード・シャノンのによって導入された。図は、事象 A_1 または A_2 が起った場合に、事象 B_1 または B_2 が生起する場合の確率の流

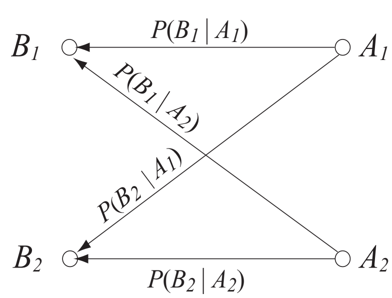

図 1.3: シャノン線図

れを表す。事象 A_1 が起る確率を $P(A_1)$ とすると、A_1 から B_1 への確率の流れは、$P(B_1|A_1)P(A_1)$ で与えられる。同様に、事象 A_2 の生起確率を $P(A_2)$ とすると、A_2 から B_1 への確率の流れは、$P(B_1|A_2)P(A_2)$ で与えられる。二つを足すと、確率 $P(B_1)$ が得られる。B_2 の場合についても同

様な解釈ができる。

$$\begin{aligned} P(B_1) &= P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2) \\ P(B_2) &= P(B_2|A_1)P(A_1) + P(B_2|A_2)P(A_2) \end{aligned}$$

このうちで、A_1 から流れ込んだ量 $P(B_1|A_1)P(A_1)$ の割合が、B_1 が起ったと分った場合にそれが A_1 に起因する確率 $P(A_1|B_1)$ に他ならない。

いま、0 と 1 を発信する情報源があるとしよう。この情報源が 0 を発信する確率を 0.6、1 を発信する確率を 0.4 とする。この信号を正確に受信する確率を、0.9、誤って受信する確率を 0.1 とする。受信信号 0 を受け取った場合に、情報源の入力信号が 0 または 1 であった確率はいくらだろうか。

入力信号が 0，1 の確率を

$$P(0_{in}) = 0.6, \quad P(1_{in}) = 0.4$$

と表わそう。それに対して、受信信号が 0、1 である確率を、それぞれ、$P(0_{out})$、$P(1_{out})$ と書こう。入力信号が正しく伝わる確率は、

$$P(0_{out}|0_{in}) = P(1_{out}|1_{in}) = 0.9$$

で、誤って伝わる確率は、

$$P(0_{out}|1_{in}) = P(1_{out}|0_{in}) = 0.1$$

と表わされる。0 または 1 を受信する確率は、それぞれ、

$$\begin{aligned} P(0_{out}) &= P(0_{out}|0_{in})P(0_{in}) + P(0_{out}|1_{in})P(1_{in}) \\ &= 0.9 \times 0.6 + 0.1 \times 0.4 = 0.58 \\ P(1_{out}) &= P(1_{out}|0_{in})P(0_{in}) + P(1_{out}|1_{in})P(1_{in}) \\ &= 0.1 \times 0.6 + 0.9 \times 0.4 = 0.42 \end{aligned}$$

と求まる。受信信号 0 を受け取った場合に、入力信号が 0 または 1 であった確率は、それぞれ、

$$\begin{aligned} P(0_{in}|0_{out}) &= \frac{P(0_{out}|0_{in})P(0_{in})}{P(0_{out})} \\ &= \frac{0.54}{0.58} \simeq 0.931 \\ P(1_{in}|0_{out}) &= \frac{P(0_{out}|1_{in})P(1_{in})}{P(0_{out})} \\ &= \frac{0.04}{0.58} \simeq 0.069 \end{aligned}$$

となる。$P(0_{in})$、$P(1_{in})$ は、結果とは無関係に事前に与えられている確率なので、事前確率と呼ばれる。一方、$P(0_{in}|0_{out})$、$P(1_{in}|0_{out})$ などは、結果 0_{out} が生じた後、すなわち、事象が起った後にその原因が 0_{in} だったのか 1_{in} だったのかを追求する確率なので、事後確率と呼ばれる。

1.4 確率変数、確率分布

サイコロを振った時の目の出方は 6 通りあり、これを、

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

と書いた。この意味をもう少し正確に表現すると、サイコロを振って1の目が出る事象を1と表したのである。他の場合 $(2,\ 3, \cdots,\ 6)$ についても同様である。このように、事象を数値で表すと何かと便利である。そのためには、事象に数値を対応させる規則を定める必要がある。この規則のことを**確率変数**と言う。確率変数とは実数の値をとる関数 (集合論で謂う所の写像) であることに注意しよう。確率変数における変数の使い方は、いわゆる変数の使い方とは異なっている。確率変数は、通常、大文字 X で表す。先ほどのサイコロの例の場合には、

$$x = X(\text{サイコロを振って } x \text{ の目が出る事象})$$

となる。ここで、左辺の小文字 x は1から6までの自然数の値をとる **変数** で、これを **確率変数** と言う場合もある。

ここらあたりは少々込み入っているので、整理してみよう。普通の意味での関数とは、ある数 x を与えると、別の数 y が得られるものであって、文字 f などを用いて表わし、

$$y = f(x)$$

と書く。x、y は、数値を代表する文字で、これらを変数と言う。一方、確率論においては、ある根源事象 ω に数値 x を対応させる式を、

$$x = X(\omega)$$

と表わし、関数 X を確率変数と言う。また、このようにして得られた変数 x も確率変数と呼ばれる。

別の例で考えてみよう。硬貨を放り投げた場合には、表が出るか裏が出るかのいずれかである。数学的に取扱い易くするためには、表と裏を数字で表すのが便利である。どんな数字を対応させるかは、まったく自由である。確率変数 X の値を、表が出た場合には 1、裏が出た場合には 0 と決めよう。これを数式を用いて表すと、

$$\begin{aligned} 1 &= X(\text{硬貨の表が出る事象}) \\ 0 &= X(\text{硬貨の裏が出る事象}) \end{aligned}$$

確率変数が、サイコロを振ったり硬貨を放ったりした場合のように、とびとびの値をとるものを**離散型**と言い、ルーレットを回した場合のように連続した値をとるものを**連続型**であるという。

確率変数 X が値 x をとる確率を $P(x)$ と表すことにすると、$P(x)$ は変数 x の関数である。そこで、これをを**確率関数**と言う。確率関数は、また、**確率分布**とも呼ばれる。 例えば、サイコロが理想的な場合には、確率分布は、

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

となる。

1.5 平均と分散

確率分布が与えられれば、確率規則は完全に決まってしまう。もはや、すべきことは何も無いと思われるかも知れない。しかし、確率のふるまいを大づかみに知りたい場合には、分布を特徴づける簡単な量があると便利である。それが平均と分散である。

平均

確率変数 x の値は、場合毎に異なっている。その典型的な値について知りたい場合に用いられるのが**平均** (mean) である。平均とは、不ぞろいなものを平らに均 (なら) すことであるが、その均し方には色々な方法がある。その中でも、よく用いられるのは、x に $P(x)$ の重みをつけて足したものである。これを、mean(平均) の頭文字 m に対応するギリシア文字 μ で表すことにすると、

$$\mu = \sum_x xP(x) \tag{1.2}$$

となる。これは、また、**期待値** (expectation) と呼ばれることもある。

例えば、理想的なサイコロを一回振る場合、1 から 6 までの目の出る確率は、各々1/6 であるので、目の平均は、

$$\mu = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2} = 3.5$$

である。

分散

平均とならんで重要なものは、確率変数の平均からの散らばり具合を表す量で、分散 (variance) と呼ばれている。これは σ^2 と表され、確率変数 x の平均 μ からのずれの 2 乗に確率 $P(x)$ の重みをつけて足したもの、すなわち、

$$\sigma^2 = \sum_x (x-\mu)^2 P(x) \tag{1.3}$$

で定義される。ここで、$(x-\mu)$ ではなく、$(x-\mu)^2$ に $P(x)$ を掛けていることに注意しよう。確率変数 x の平均 μ からのずれは、正の場合も負の場合のあるので、$(x-\mu)$ に確率 $P(x)$ の重みをつけて足したものは、0 となってしまうからである。分散は、$(x-\mu)$ の 2 乗について計算しているので、実際の散らばり具合を知るためには、その平方根をとる必要がある。これを**標準偏差** (standard deviation) と呼び、s に対応するギリシア文字 σ で表す。すなわち、

$$\sigma = \sqrt{\sigma^2}.$$

先ほどのサイコロの例で、分散を求めてみると、

$$\sigma^2 = \frac{1}{6}\sum_{x=1}^{6}\left(x-\frac{7}{2}\right)^2 = \frac{35}{12} \simeq 2.9$$

したがって、標準偏差は、

$$\sigma = \sqrt{\frac{35}{12}} \simeq 1.7$$

となる。サイコロの目の広がりは、およそ、$\mu-\sigma$ から $\mu+\sigma$ の範囲である。

分散の計算には、(1.3) を変形して得られる、もっと便利な公式がある。(1.3) 式における $(x-\mu)^2$ を展開すると、

$$\begin{aligned}
\sigma^2 &= \sum_x (x-\mu)^2 P(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) P(x) \\
&= \sum_x x^2 P(x) - 2\mu \sum_x xP(x) + \mu^2 \sum_x P(x)
\end{aligned}$$

ここで、第2項の $\sum_x xP(x)$ は、平均の定義より μ である。第3項の $\sum_x P(x)$ は、確率の総和だから1である。これらを代入して整理すると、

$$\sigma^2 = \sum_x x^2 P(x) - \mu^2 \tag{1.4}$$

を得る。

上の公式を用い、先ほどのサイコロの例について、もう一度分散を計算してみよう。

$$\begin{aligned}
\sigma^2 &= \frac{1}{6}(1^2 + 2\ 2 + 3^3 + 4^2 + 5^2 + 6^2) - \left(\frac{7}{2}\right)^2 \\
&= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) - \left(\frac{7}{2}\right)^2 \\
&= \frac{91}{6} - \frac{49}{4} \\
&= \frac{35}{12}
\end{aligned}$$

確かに、二つの計算結果は一致していることが分る。

第2章　離散分布

2.1　ベルヌーイ分布

もっとも簡単な確率分布は、「成功か失敗か」、「表か裏か」、「丁か半か」、というように結果が2つしかない場合である。この分布は、ヤコブ・ベルヌーイにちなんで**ベルヌーイ分布**と呼ばれる。ベルヌーイ家は、17世紀末から18世紀にかけて、およそ1世紀の間に8人もの優れた数学者を輩出した一族である。その中で、ヤコブ・ベルヌーイは、確率論において重要な業績を挙げている。彼の名を不朽なものにしたのは、後に出てくる大数の法則の発見によってである。

ここでは、成功か失敗かのいずれかの事象が起きる場合について考えよう。確率変数 x を導入し、成功の場合を $x = 1$、失敗の場合を $x = 0$ とし、その各々の確率を、p、$q = 1 - p$ としよう。

$$\begin{aligned} P(1) &= p, \\ P(0) &= q = 1 - p \end{aligned}$$

この分布における確率変数 x の平均 μ は、

$$\mu = 1 \times p + 0 \times q = p$$

分散 σ^2 は、

$$\sigma^2 = (1 - p)^2 + (0 - p)^2 = pq$$

となる。

上記のような試行を独立に何回も続けて行ったものを、**ベルヌーイ試行** (Berunoulli trials) と言う。これから重要な確率分布のいくつかが導かれる。

2.2　幾何分布

2.2.1　幾何分布の平均と分散

分布関数が

$$p(x) = pq^{x-1} \quad \text{ただし、} p \text{ は、0 と 1 の間の定数で、} q = 1 - p$$

で与えられる幾何分布の平均と分散は、以下の式で与えられる。

$$\text{平均} \quad : \quad \mu = \sum_{x=1}^{\infty} xp(x) = p\left[\sum_{x=1}^{\infty} xq^{x-1}\right] \tag{2.1}$$

$$\text{分散} \quad : \quad \sigma^2 = \sum_{x=1}^{\infty} (x - \mu)^2 p(x) = p\left[\sum_{x=1}^{\infty} x^2 q^{x-1}\right] - \mu^2 \tag{2.2}$$

ここで、$S_0(q)$、 $S_1(q)$、 $S_2(q)$ を次のように定義する。

$$\begin{aligned}
S_0(q) &= \sum_{x=1}^{\infty} q^{x-1} = 1 + q + q^2 + q^3 + q^4 + q^5 + q^6 + q^7 + q^8 + \dots \\
S_1(q) &= \sum_{x=1}^{\infty} x q^{x-1} = 1 + 2q + 3q^2 + 4q^3 + 5q^4 + 6q^5 + 7q^6 + 8q^7 + 9q^8 + \dots \\
S_2(q) &= \sum_{x=1}^{\infty} x^2 q^{x-1} = 1 + 4q + 9q^2 + 16q^3 + 25q^4 + 36q^5 + 49q^6 + 64q^7 + 81q^8 + \dots
\end{aligned}$$

すると、平均と分散は、 $S_1(q)$、 $S_2(q)$ を用いて以下のように書ける。

$$\text{平均} \quad : \quad \mu = \sum_{x=1}^{\infty} x p(x) = p S_1(q) \tag{2.3}$$

$$\text{分散} \quad : \quad \sigma^2 = \sum_{x=1}^{\infty} (x-\mu)^2 p(x) = p S_2(q) - \mu^2 \tag{2.4}$$

で与えられる。従って、平均と分散を求めるには、 無限級数 $S_1(q)$、$S_2(q)$ が求まれば良い。

2.2.2 無限等比級数 $S_0(q)$ の求め方

先ず、$S_0(q)$ を求める。そのためには、$S_0(q)$ を q 倍し、$S_0(q)$ から引けば良い。

$$\begin{array}{rrclllllllllll}
 & S_0(q) & = & 1 & + & q & + & q^2 & + & q^3 & + & q^4 & + & \cdots \\
-) & qS_0(q) & = & & & q & + & q^2 & + & q^3 & + & q^4 & + & \cdots \\
\hline
 & (1-q)S_0(q) & = & 1 & & & & & & & & & &
\end{array}$$

ゆえに、

$$S_0(q) = \frac{1}{1-q} \tag{2.5}$$

を得る。

2.2.3 平均の求め方

平均 μ を求めるためには、無限級数の和 $S_1(q)$ を求めれば良い。$S_0(q)$ を求めた方法と同様に、$S_1(q)$ を q 倍し、$S_1(q)$ から引いてみよう。

$$\begin{array}{rrclllllllllll}
 & S_1(q) & = & 1 & + & 2q & + & 3q^2 & + & 4q^3 & + & 5q^4 & + & \cdots \\
-) & qS_1(q) & = & & & q & + & 2q^2 & + & 3q^3 & + & 4q^4 & + & \cdots \\
\hline
 & (1-q)S_1(q) & = & 1 & + & q & + & q^2 & + & q^3 & + & q^4 & + & \cdots
\end{array}$$

引き算した結果、$(1-q)S_1(q)$ の右辺は、無限等比級数の和 $S_0(q)$ であることが分かる。すなわち、

$$\begin{aligned}
(1-q)S_1(q) &= S_0(q) \\
&= \frac{1}{1-q}
\end{aligned}$$

ゆえに、

$$S_1(q) = \frac{1}{(1-q)^2} \tag{2.6}$$

を得る。ここで、関係式 $q=1-p$ を用いると、

$$S_1(q)=\frac{1}{p^2} \tag{2.7}$$

を得る。(2.7) 式を (2.3) 式に代入して、平均を表す最終的な結果、

$$\mu=\frac{1}{p} \tag{2.8}$$

を得る。

2.2.4 分散の求め方

分散 σ を求めるためには、無限級数の和 $S_2(q)$ を求めれば良い。前回と同様に、 $S_2(q)$ を q 倍し、$S_2(q)$ から引いてみよう。

$$\begin{array}{lrcccccccccccc}
 & S_2(q) & = & 1 & + & 4q & + & 9q^2 & + & 16q^3 & + & 25q^4 & + & \cdots \\
-) & qS_2(q) & = & & & q & + & 4q^2 & + & 9q^3 & + & 16q^4 & + & \cdots \\
\hline
 & (1-q)S_2(q) & = & 1 & + & 3q & + & 5q^2 & + & 7q^3 & + & 9q^4 & + & \cdots
\end{array}$$

引き算の結果の $(1-q)S_2(q)$ の右辺は、$S_1(q)$ を 2 倍して $S_0(q)$ を引いても得られる。

$$\begin{array}{lrcccccccccccc}
 & 2S_1(q) & = & 2 & + & 4q & + & 6q^2 & + & 8q^3 & + & 10q^4 & + & \cdots \\
-) & S_0(q) & = & 1 & + & q & + & q^2 & + & q^3 & + & q^4 & + & \cdots \\
\hline
 & 2S_1(q)-S_0(q) & = & 1 & + & 3q & + & 5q^2 & + & 7q^3 & + & 9q^4 & + & \cdots
\end{array}$$

ゆえに、

$$(1-q)S_2(q)=2S_1(q)-S_0(q)$$

が成り立つ。これを、$S_2(q)$ について解くと、

$$S_2(q)=\frac{2S_1(q)-S_0(q)}{1-q} \tag{2.9}$$

(2.9) に (2.6) と (2.5) 代入して、

$$S_2(q)=\frac{1+q}{(1-q)^3} \tag{2.10}$$

を得る。変数 q の代わりに p を用いて表すと、

$$S_2(q)=\frac{2-p}{p^3} \tag{2.11}$$

となる。(2.4) 式に、(2.11) 式と (2.8) 式を代入して、幾何分布の分散の最終的な形、

$$\sigma^2=\frac{1-p}{p^2} \tag{2.12}$$

を得る。

2.3 二項分布

2.3.1 定義

一回の試行で成功する確率を p、失敗する確率を $q = 1 - p$ とする。n 回の試行で、x 回成功する確率は、

$$p(x) = {}_nC_x p^x q^{n-x}, \quad (0 \leq x \leq n)$$

で与えられる。このような分布を二項分布と言い、B(n, p) で表わす。

確率の総和

$$\begin{aligned}\sum_{x=0}^{n} p(x) &= \sum_{x=0}^{n} {}_nC_x p^x q^{n-x} \\ &= (p+q)^n \\ &= 1\end{aligned}$$

2.3.2 二項分布の平均と分散

平均

$$\begin{aligned}\mu &= \sum_{x=0}^{n} xp(x) \\ &= \sum_{x=1}^{n} x {}_nC_x p^x q^{n-x}\end{aligned}$$

ところで、

$$x {}_nC_x = x\frac{n!}{x!(n-x)!} = \frac{n(n-1)!}{(x-1)!\big((n-1)-(x-1)\big)!} = n {}_{n-1}C_{x-1}$$

だから、

$$\begin{aligned}\sum_{x=1}^{n} x {}_nC_x p^x q^{n-x} &= np\sum_{x=1}^{n} {}_{n-1}C_{x-1} p^{x-1} q^{(n-1)-(x-1)} \\ &= np\sum_{x'=0}^{n-1} {}_{n-1}C_{x'} p^{x'} q^{(n-1)-x'} \\ &= np(p+q)^{n-1} \\ &= np \end{aligned} \tag{2.13}$$

ここで、最後の式の変形では、$p+q=1$ であることを用いた。したがって、二項分布の平均は、

$$\mu = np$$

となる。

分散

$$\sigma^2 = \sum_{x=0}^{n} x^2 p(x) - \mu^2$$

だから、第 1 項の $\sum_{x=0}^{n} x^2 p(x)$ が求まれば良い。

$$\begin{aligned} \sum_{x=0}^{n} x^2 p(x) &= \sum_{x=0}^{n} [x(x-1)+x]_n C_x p^x q^{n-x} \\ &= \sum_{x=2}^{n} x(x-1)_n C_x p^x q^{n-x} + \sum_{x=1}^{n} x_n C_x p^x q^{n-x} \end{aligned} \tag{2.14}$$

ここで (2.14) 式の第 1 項は、

$$x(x-1)_n C_x = n(n-1)_{n-2} C_{x-2}$$

と書けることを使うと、

$$\begin{aligned} \sum_{x=2}^{n} x(x-1)_n C_x p^x q^{n-x} &= n(n-1)p^2 \sum_{x'=0}^{n-2} {}_{n-2}C_{x'} p^{x'} q^{(n-2)-x'} \\ &= n(n-1)p^2 (p+q)^{n-2} \\ &= n(n-1)p^2 \end{aligned}$$

$$\begin{aligned} \sum_{x=0}^{n} x(x-1)p(x) &= n(n-1)p^2 \sum_{x=2}^{n-2} {}_{n-2}C_{x-2} p^{x-2} q^{(n-2)-(x-2)} \\ &= n(n-1)p^2 (p+q)^{n-2} \\ &= n(n-1)p^2 \end{aligned}$$

また、(2.14) 式の第 2 項は x の平均で、すでに (2.13) 式で求めたように、np である。以上のことから、

$$\sigma^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq$$

と求まる。

第3章　正規分布

3.1　正規分布の性質

確率変数が連続的な値を取るものを、連続分布と言う。その中で最も重要なのが、**正規分布** (normal distribution) である。正規分布とは、ずいぶん堅苦しい言い方だが、normal とは、標準的なとか平均的などの意味である。正規分布とは、いろんな所にあまねく存在する分布である。ある一つの物を何度も測定すると、その測定値は、ほとんど同じであるが、皆少しずつ異なっている。この測定誤差の分布が正規分布である。そんなわけで、この分布は、別名、**誤差分布**とも呼ばれる。試験の成績分布も、受験者の数が多い場合には、正規分布 (誤差分布) に近くなる。試験の誤りと測定の誤りの原因は、全く異なっているにもかかわらず、似たような分布になるのである。

正規分布は、最初、ド・モアブルにより二項分布の極限として導入された。そして、この分布を詳しく調べたのが、ドイツの数学者カール・フリードリッヒ・ガウス (1777〜1855) であったので、ガウス分布とも言う。確率変数 x の取り得る範囲は、マイナス無限大からプラス無限大までの全領域である。平均が μ、標準偏差が σ で与えられる正規分布の確率密度関数 $f(x)$ は、

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

で与えられる。これを $N(\mu, \sigma^2)$ と表す。

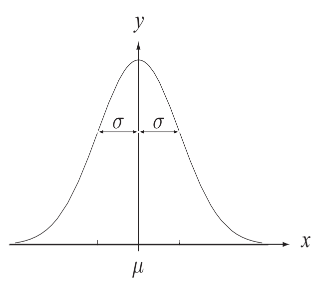

図 3.1: 正規分布

正規分布の特徴は、確率密度関数が、平均値 $x = \mu$ に対して対象なことである。

正規分布のような連続分布において、確率変数 x が、ある特定の値を取る確率というのは意味を持たない。なぜならばその確率は常にゼロだからである。連続分布において意味があるのは、確率変数 x がある値の範囲を取る場合である。すべての確率についての和は、

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} \, dx = 1$$

と表される。
証明

$$I = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2}\ dx$$

と置く。ここで、$u = (x-\mu)/\sqrt{\sigma}$ と変数変換をすると、

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2}\ du$$

となる。ここで、積分 I の代わりに、I^2 を考える。

$$I^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2}\ du \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-v^2/2}\ dv = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(u^2+v^2)/2}\ du\ dv$$

ここで、$u = r\cos\theta,\ \ v = r\sin\theta$ とおいて、極座標 $(r,\ \theta)$ に移行する。
関係式 $u^2+v^2 = r^2,\ \ dudv = 2\pi rdr$ を用いて、書きなおすと、

$$I^2 = \int_0^{+\infty} e^{-r^2/2}\ rdr = \left[-e^{-r^2/2}\right]_0^{+\infty} = 1$$

ゆえに、

$$I = 1$$

証明終わり。

確率変数 x が、a から b までの範囲を取る確率 $P(a<x<b)$ は、確率密度関数 $f(x)$ を a から b まで積分することにより得られる。

$$P(a<x<b) = \int_a^b f(x)\ dx \tag{3.1}$$

この値を解析的に求めることはできないので、数値積分を行うしかない。しかし、すべての正規分布は、変数変換により、平均が 0、分散が 1 の正規分布に変換することができる。これを**標準正規分布**と言う。標準正規分布についての積分表が与えられていれば、すべての正規分布についての確率が分ることになる。

——— ここは、飛ばしても良い。 ———-

平均が μ、標準偏差が σ^2 の正規分布正規分布関数 $f(x)dx = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx$ を標準正規分布に変換するには、ディラックのデルタ関数 $\delta(z - \frac{x-\mu}{\sigma})$ を掛けて、変数 x について積分すればよい。ここで、$\delta(z-\frac{x-\mu}{\sigma}) = \sigma\delta(\sigma z - x + \mu)$ と書けるので、

$$\begin{aligned}
\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \delta\Big(z - \frac{x-\mu}{\sigma}\Big)\, dx &= \int_{-\infty}^{+\infty} \frac{\sigma}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \delta(\sigma z - x + \mu)\, dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}
\end{aligned}$$

となるので、確率変数 z は、平均が 0、分散が 1 の正規分布に従う。

全ての正規分布は、標準正規分布に変換できるので、標準正規分布の計算ができれば、すべての正規分布の計算ができることになる。例えば、正規分布 $N(\mu,\ \sigma^2)$ において、確率変数 x が、a と b の間の値をとる確率 $P(a \leqq x \leqq b)$ は、標準正規分布を用いて、

$$P(a \leqq x \leqq b) = P_{標準}\left(\frac{a-\mu}{\sigma} \leqq z \leqq \frac{b-\mu}{\sigma}\right)$$

表される。

さらに、標準正規分布は、$z=0$(y 軸) に関して対称だから、巻末の正規分布表にあるような

$$P_{標準}(K_P \leqq z) = \frac{1}{\sqrt{2\pi}} \int_{K_P}^{\infty} e^{-z^2/2}\ dz$$

の値が分れば、すべての正規分布の値は計算できることになる。

正規分布表の見方

正規分布表においては、K_P の小数点以下 1 桁目までが縦軸に、2 桁目が横軸に書かれている。例えば、$P_{標準}(1.23 \leqq z)$ の値を知りたい場合には、K_P の欄を下に見ていくと、1.2 と書かれた箇所がある。次に、K_P の欄を横にみていくと 0.03 と書かれた場所がある。そこで、1.2 のある行と 0.03 のある列の交差点を見ると、0.10935 と書かれた数字がある。これが、$P_{標準}(1.23 \leqq z)$ の値である。すなわち、

$$P_{標準}(1.23 \leqq z) = 0.10935$$

である。

例題

確率変数 z が、$N(0,1)$ にしたがうときに、$P_{標準}(-0.67 \leqq z \leqq 1.23)$ を求めてみよう。

$$\begin{aligned} P_{標準}(-0.67 \leqq z \leqq 1.23) &= 1 - P_{標準}(z \leqq -0.67) - P_{標準}(1.23 \leqq z) \\ &= 1 - P_{標準}(0.67 \leqq z) - P_{標準}(1.23 \leqq z) \\ &= 1 - 0.25143 - 0.10935 \\ &= 0.63922 \end{aligned}$$

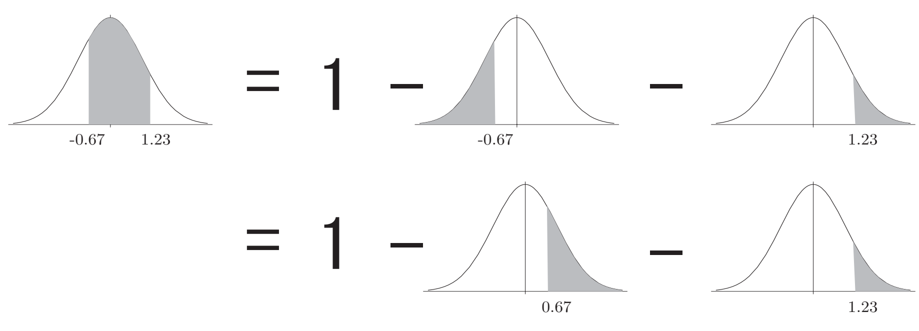

図 3.2: 任意の区間での正規分布の確率は、$P_{標準}(K_P \leqq z)$ を用いて計算できる。

正規分布において特に重要な確率の値は、平均値からのずれが標準偏差の 1 倍、2 倍、3 倍などの場合である。

$$P(\ -\sigma\ \leqq \mu \leqq\ \sigma\ )\ \ =\ \ P_{標準}(-1 \leqq z \leqq 1) = 0.683$$

$$
\begin{aligned}
P(-2\sigma \leqq \mu \leqq 2\sigma) &= P_{\text{標準}}(-2 \leqq z \leqq 2) = 0.955 \\
P(-3\sigma \leqq \mu \leqq 3\sigma) &= P_{\text{標準}}(-3 \leqq z \leqq 3) = 0.997
\end{aligned}
$$

一番目の式からは、平均からのずれが標準偏差以内に収まる確率は、およそ 68 %であることがわかる。標準偏差以上および以下のずれの確率は、各々16 %ほどである。

3.2 偏差値

標準正規分布では、平均を 0 に、標準偏差を 1 にとる。しかし、この方法が、いつも分り易いとは限らない。標準化が必要な場合でも、平均の値や標準偏差の値は、ケースバイケースで違うのである。たとえ 100 点満点の試験であっても、試験毎に平均点や標準偏差が異なるので、違う試験の結果同士を直接に比較することはできない。二つを比較するためには、標準化をする必要がある。例えば、最初は 65 点だったのが、2 回目の試験で 73 点だったとしよう。この場合に、2 回目の試験のほうが成績が良かったといえるだろうか。2 回目の試験が 1 回目に比べて単に易しかっただけなのかも知れないので、実は何も言えないのである。このような場合の比較には、平均を 0 標準偏差を 1 とするよりも、平均が 50 で標準偏差が 10 の方が直感的で分り易いことは明らかである。このようにして求めた得点を学力偏差値と言う。もともとは、高校の理科の先生が、受験生の不安を和らげるために、成績の客観的に評価をめざして考え出したものだそうである。しかし、それ以降、学校教育が偏差値一辺倒となってしまったために、偏差値の評判は、はなはだ悪いものとなってしまった。ここでは、偏差値の功罪については立ち入らずに、どんなものなのかだけを説明しよう。

いま、100 点満点の試験で、平均値が μ、標準偏差が σ だったとしよう。この試験における得点を x 点とする。これを、平均を 50、標準偏差を 10 に調整した場合の得点 (偏差値)T は、

$$
T = \frac{x-\mu}{\sigma} \times 10 + 50
$$

で与えられる。 このことは、$x = \mu$ とおけば、$T = 50$ となること、また、$x = \mu + \sigma$ とおけば $T = 60$ に、$x = \mu - \sigma$ とおけば $T = 40$ になることかも容易に分る。

例題

A 君の前期試験の成績は 67 点、後期試験の成績は 71 点だった。前期試験の平均点は 54 点、標準偏差は 14 点であった。一方、後期試験の平均点は 52 点、標準偏差は 26 点だった。A 君の成績は上がったと言えるだろうか。

	得点	平均	標準偏差
前期	67	54	14
後期	71	52	26

平均点は、54 点から 52 点に、わずかだが下がっている。一方、A 君の得点は、67 点から 71 点に上昇している。単純に考えると。A 君の成績は上がったように見える。ところが、前期試験と後期試験の偏差値を計算してみると、

$$
\begin{aligned}
T_{\text{前期}} &= \frac{67-54}{14} \times 10 + 50 \simeq 59.3 \\
T_{\text{後期}} &= \frac{71-52}{26} \times 10 + 50 \simeq 57.3
\end{aligned}
$$

となり、後期試験の方が低くなっている。これは、前期試験の得点分布が、後期試験に比べて、平均点の周りに集中していた結果である。標準偏差についての補正を行うと、後期試験の結果の方が悪いことが分る。

3.3 正規分布の線形結合

x が正規分布 $N(\mu, \sigma^2)$ に従っているとき、その線形変換 $y = ax + b$ は正規分布 $N(a\mu + b, a^2\sigma^2)$ に従う。

証明

確率変数 x は、平均 μ、分散 σ^2 の正規分布 $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ に従うとする。確率変数が $y = ax + b$ である分布を $F(y)$ と書くことにすると、

$$\begin{aligned} F(y) &= \int_{-\infty}^{+\infty} \delta(ax + b - y) f(x)\, dx \\ &= \frac{1}{a} f\left(\frac{y-b}{a}\right) dx \\ &= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\left[-\frac{\big(y - (a\mu + b)\big)^2}{2a^2\sigma^2}\right] dy \end{aligned}$$

これは、確率変数 y が平均 $a\mu + b$、分散 $a^2\sigma^2$ の正規分布 $N(a\mu + b, a^2\sigma^2)$ に従うことを表している。

3.4 正規分布の一次結合

確率変数 x_1、x_2 が互いに独立で、それぞれ正規分布 $N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$ に従っているとき、確率変数 $x = x_1 + x_2$ は正規分布 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ に従う。

それでは、２つの正規分布関数をたたみこんだ関数 $f(x)$ を計算しよう。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \int \exp\Big[-\frac{(x - x_2 - \mu_1)^2}{2\sigma_1^2} - \frac{(x_2 - \mu_2)^2}{2\sigma_2^2}\Big]\, dx_2$$

指数関数の肩の部分を整理すると、

$$\begin{aligned} &-\frac{(x - x_2 - \mu_1)^2}{2\sigma_1^2} - \frac{(x_2 - \mu_2)^2}{2\sigma_2^2} \\ &= -\frac{1}{2\sigma_1^2\sigma_2^2}\left\{(\sigma_1^2 + \sigma_2^2)\left[x_2 - \frac{\sigma_2^2(x - \mu_1) + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right]^2 + \frac{\sigma_1^2\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)}(x - \mu_1 - \mu_2)^2\right\} \end{aligned}$$

となる。ここで、x_2 についての積分は、

$$\int_{-\infty}^{+\infty} \exp\left\{-\frac{(\sigma_1^2 + \sigma_2^2)}{2\sigma_1^2\sigma_2^2}\left[x_2 - \frac{\sigma_2^2(x - \mu_1) + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right]^2\right\} dx_2 = \sqrt{\frac{2\pi\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}$$

となるので、

$$f(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left[-\frac{(x - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right]$$

となる。これは、平均が $\mu_1 + \mu_2$、分散が $\sigma_1^2 + \sigma_2^2$ の正規分布を表す。

上記の結果と、先に示した正規分布の線形変換の結果を合わせると、一般的に次のことが言える。

確率変数 $x_1,\ x_2,\ x_3,\ \cdots,\ x_n$ が互いに独立で、それらの分布が正規分布 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$, $N(\mu_3, \sigma_3^2), \cdots,\ N(\mu_n, \sigma_n^2)$ であるとき、

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots + a_n x_n$$

は、

$$\text{平均} \qquad \mu_y = a_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + \cdots + a_n\mu_n, \tag{3.2}$$

$$\text{分散} \qquad \sigma_y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2 + \cdots + a_n^2\sigma_n^2 \tag{3.3}$$

の正規分布 $N(\mu_y, \sigma_y^2)$ に従う。