

この文書は西郷達彦・有本彰雄『R による極値統計学』（オーム社、2020）の追加説明である。一部に高校レベルを超える大学教養の数学が使われているため、その説明をする。さらに詳しくは微分積分と線形代数の教科書を見て欲しい。ただしそれらで定義が以下に述べるものと同値な別の言い方になっていることがある。

1 記号

記号	説明
$\sum_k a_k$	$\sum_{k=1}^n a_k$ または $\sum_{k=1}^{\infty} a_k$ の省略形。
$\sum_{x_k \leq x} p(x_k)$	$x_k \leq x$ をみたす x_k について $p(x_k)$ の和を取る。
$\binom{n}{x}$	$\frac{n(n-1)\dots(n-x+1)}{x!}$. n が自然数の場合 ${}_nC_x$ と同じ。
$A \doteq B$	A と B の値が近くほぼ同じ値であることを表す。
$A_n \approx B_n, A_n \sim B_n$	$\lim_{n \rightarrow \infty} \frac{A_n}{B_n} = 1$ となること。 $n \approx n + \log n$ のように差が無限になることもある。
$A := B$	A を B で定義すること。
$\mathbf{N}, \mathbf{Q}, \mathbf{R}, \mathbf{C}$	それぞれ、自然数・有理数・実数・複素数の集合。

2 行列

2.1 定義

長方形状に数値を並べてカッコでくくったものを行列という。多次元のデータを数の組であるベクトルで表し、これに伴って行列が必要となる。

行列は次のようにアルファベットの太文字で表す。

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

この行列の中で $(a_{i1}, a_{i2}, \dots, a_{in})$, $1 \leq i \leq m$ となる横ベクトルを第 i 行といい、

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}, 1 \leq j \leq n$$

となる縦ベクトルを第 j 列といい、行列 A を (m, n) 型行列という。行列において第 i 行で第 j 列に当たる要素を (i, j) 成分といい、これを用いて上の行列を $A = (a_{ij})$ と略記する。

縦ベクトルを表すのに印刷スペースを節約するため、後に説明する転置の記号を使い $\mathbf{a}_j^T = (a_{1j}, a_{2j}, \dots, a_{mj})$ と表すことがある。

2.2 演算

同じ (m, n) 型の行列 A, B があるとき、その和と差が次のように定義される。

$$A + B = (a_{ij} + b_{ij}), A - B = (a_{ij} - b_{ij})$$

例えば

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} + \begin{pmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 7 & 5 \\ 10 & 15 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} - \begin{pmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ -1 & 3 \\ 0 & -3 \end{pmatrix}$$

となる。

(m, n) 型行列 $A = (a_{ij})$ と実数 k について、スカラー倍を $kA = (ka_{ij})$ なる (m, n) 型行列として定義する。スカラー倍により $(-1)A$ と定義される行列を $-A$ と表す。

(ℓ, m) 型行列 A と (m, n) 型行列 B について積 AB が次のように定義され (ℓ, n) 型行列となる。

$$AB = \left(\sum_{k=1}^m a_{ik} b_{kj} \right)$$

すなわち左側の行列 A の第 i 行と右側の行列 B の第 j 列は次元が同じ m であり、その内積を積 AB の (i, j) 成分とする。左側の行列の列数と右側の行列の行数が等しいときのみ定義されるため、 AB が定義されても BA が定義されるときは限らないし、もし BA が定義されたとしても $AB = BA$ となるとは限らず、交換法則は成り立たない。しかし結合法則 $(AB)C = A(BC)$ は積が定義できれば成り立つ。 (n, n) 型行列 A について累乗を $A^2 = AA, A^3 = A^2A$ などと表す。

(m, n) 型行列 A についてその第 i 行を第 i 列とする新たな (n, m) 型行列を A の転置行列といい、 A^T と表す。すなわち A^T の (i, j) 成分は a_{ji} である。

2.3 特殊な行列

すべての成分が 0 である行列を零行列といい、 O で表す。

(m, n) 型行列において $m = n$ であるとき n 次正方行列という。

n 次正方行列 $A = (a_{ij})$ において対角成分すなわち a_{ii} , $1 \leq i \leq n$ がすべて 1 であり、それ以外の成分がすべて 0 であるとき、 A を単位行列といい、 E または I で表す。単位行列 I に対し、行列の積 BI が定義されるとき、 $BI = B$ となる。

n 次正方行列 A について $AB = BA = I$ となる n 次正方行列 B が存在するとき、 A は正則といい、 B を A の逆行列という。さらに $B = A^{-1}$ と表す。

n 次正方行列 A が $A^T = A$ を満たすとき対称行列という。

n 次正方対称行列 A についてどんな n 次元縦ベクトル \mathbf{v} をとっても

$$\mathbf{v}^T A \mathbf{v} \geq 0$$

となるとき、 A は非負定値であるという。スカラーの非負に当たる。

確率ベクトル $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ があるとき、その平均を $E(\mathbf{X}) = (E(X_1), E(X_2), \dots, E(X_n))^T$ と定義する。このとき $(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^T$ は n 次正方行列であり、その (i, j) 成分の平均を (i, j) 成分とする行列を \mathbf{X} の共分散行列という。共分散行列の (i, j) 成分は X_i と X_j の共分散となっている。また共分散行列は常に対称で非負定値である。

3 偏微分と関数の最大最小

2 変数関数 $z = f(x, y)$ が最大値を取る点 $(x, y) = (a, b)$ を求めることを考える。最尤法における尤度関数の最大化などは典型的な例である。

1 変数関数 $y = g(x)$ の最大化問題では、 $g(x)$ が微分可能であれば微分 $g'(x)$ を求め、 $g'(a) = 0$ となる点 $x = a$ が最大値となる点の候補である。同様にして 2 変数関数においても最大値を求めるときに微分を考える。

3.1 偏微分

2 変数関数の微分としては全微分と偏微分があり、ここでは偏微分を用いる。偏微分は 2 変数関数で一方の変数を定数と見て他方の変数で微分するものである。定義は点 (a, b) において

$$\begin{aligned} \frac{\partial f}{\partial x}(a, b) &= f_x(a, b) := \lim_{h \rightarrow 0} \frac{f(a+h, b) - f(a, b)}{h} \\ \frac{\partial f}{\partial y}(a, b) &= f_y(a, b) := \lim_{h \rightarrow 0} \frac{f(a, b+h) - f(a, b)}{h} \end{aligned}$$

の右辺が存在する（有限確定値になる）とき左辺を (a, b) における偏微分係数という。そこで各点について偏微分係数が求まるため、点 (x, y) から偏微分係数 $f_x(x, y), f_y(x, y)$ への関数を考えることができ、これを偏微分または偏導関数という。

3.2 極値

偏導関数を用いると2変数関数の極値を求める問題を考えることができる。まず2変数のときの極値（極値統計の極値とは別物である）の定義を与える。ある $r > 0$ について (a, b) を中心とする半径 r の円の内部 $S_r(a, b) = \{(x, y) | (x - a)^2 + (y - b)^2 < r^2\}$ を (a, b) を中心とする開球という。平面 \mathbf{R}^2 上の点 (a, b) が関数 $f(x, y)$ が極大であるとは、ある $r > 0$ が存在して開球 $S_r(a, b)$ において点 (a, b) 以外のかつてな点 (x, y) で $f(a, b) > f(x, y)$ となることである。つまりごく小さくてもいいのでその点の周りで、その点での関数の値が周りよりも大きいことである。極小は逆向きの不等号が成り立つときである。

この極大となる点を求めるときには偏導関数を用いる。2変数関数 $f(x, y)$ の領域 $D \subset \mathbf{R}^2$ で偏導関数が存在し、また点 $(a, b) \in D$ で $f(x, y)$ が極大であれば、 $f_x(a, b) = f_y(a, b) = 0$ となる。

この結果を用いると、偏導関数が0になる点を求めることで極大になる点の候補を見つけることができる。また極大になる点は最大になる点の候補である。よって偏導関数を調べることで最大値を取る点の候補を見つけることができる。これは1変数関数 $y = g(x)$ において最大値を取る点を見つける方法と同様である。もとの関数の性質がよく領域が \mathbf{R}^2 全体などであれば、偏導関数を調べるだけで最大値を取る点が求められ、実際に統計学や工学の多くの問題でこの方法を使い最大・最小となる点を求めている。

2変数より変数の多い一般の多変数関数についても同様の結果が成り立つ。

4 重積分

2変数関数 $z = f(x, y)$ に関する積分を考える。グラフにすると平面上の点 (x, y) により z が決まるため、 $z = f(x, y) > 0$ であれば、ある領域 D の上側に屋根があるようなグラフとなる。この領域 D について屋根と xy 平面で挟まれた部分の体積を求める問題を考える。2次元正規分布の密度関数が与えられたときの確率を求める問題などが典型例である。

1変数関数 $y = f(x)$ であれば、区間 $[a, b]$ において $y = f(x) > 0$ のとき、この区間でグラフ $y = f(x)$ と x 軸に挟まれた部分の面積は積分で求められる。積分は定義としては区分求積法で定義され、計算としては微分の逆演算として行われる。

2変数関数についても同様にして、上の問題の体積は重積分で求められる。重積分は区分求積の考え方で定義され、計算としては微分の逆演算を x, y それぞれについて行う逐次積分で行われる。

4.1 定義

まず平面 \mathbf{R}^2 上の長方形 $[a, b] \times [c, d]$ において 2 変数関数 $f(x, y)$ の重積分を定義する。考え方としては長方形を小さな長方形に分割して、各小長方形において柱状の立体の体積を計算し、足し合わせるものである。区間 $[a, b]$ を分割し $a = x_0 < x_1 < \dots < x_m = b$ とし、また $[c, d]$ を分割し $c = y_0 < y_1 < \dots < y_n = d$ とする。さらに $\Delta x = \max_{1 \leq i \leq m} (x_i - x_{i-1})$, $\Delta y = \max_{1 \leq j \leq n} (y_j - y_{j-1})$ とおく。そこで小さい長方形 $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$, $1 \leq i \leq m, 1 \leq j \leq n$ からかってな点 $(\xi_{i,j}, \eta_{i,j})$ を選ぶ。この時分割を細かくして

$$\lim_{\Delta x, \Delta y \rightarrow 0} \sum_{i=1}^m \sum_{j=1}^n f(\xi_{i,j}, \eta_{i,j}) (x_i - x_{i-1}) (y_j - y_{j-1})$$

が分割と代表点の選び方に関わらず一定の値に収束するとき、その極限値を $f(x, y)$ の $[a, b] \times [c, d]$ における重積分といい、 $\int \int_{[a,b] \times [c,d]} f(x, y) dx dy$ と表す。

一般の有界領域 $D \subset [a, b] \times [c, d]$ においては、新たな関数 $f_1(x, y)$ を導入し、 $(x, y) \in D$ のとき $f_1(x, y) = f(x, y)$ とおき、 $x \notin D$ のとき $f_1(x, y) = 0$ とおく。この関数 $f_1(x, y)$ を $[a, b] \times [c, d]$ で重積分した値を、関数 $f(x, y)$ の領域 D における重積分と定義し、 $\int \int_D f(x, y) dx dy$ と表す。

2 変数関数 $f(x, y)$ について \mathbf{R}^2 全体のような有界でない領域すなわちどんな $[a, b] \times [c, d]$ についてその部分集合にならない領域 D についての積分を考える。領域 D に含まれる有界部分領域の列を $D_1 \subset D_2 \subset \dots \subset D_n \subset D$ とする。そこで重積分の列 $S_n = \int \int_{D_n} f(x, y) dx dy$ について、その極限 $\lim_{n \rightarrow \infty} S_n$ が部分領域の列の取り方によらず一定値に収束するとき、 D での広義積分と定義する。1 変数関数の広義積分も同様である。

2 変数を超える多変数関数についても同様である。また逐次積分についてはやや技術的であり、本文で用いないので省略する。

5 テイラー展開と近似

指数関数や対数関数あるいは三角関数などを多項式の関数で近似することを考える。

ある関数 $f(x)$ が無限級数により

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n + \dots = \sum_{k=0}^{\infty} a_k x^k$$

と表されたとする。そこで係数 $\{a_k\}$ がなんであるか知りたい。微分と無限和の順序交換が可能であれば、

$$f(0) = a_0, f'(0) = a_1, f''(0) = 2a_2, f^{(3)}(0) = 3 \cdot 2a_3, \dots, f^{(n)}(0) = n!a_n, \dots$$

となる。よって

$$a_k = \frac{1}{k!} f^{(k)}(0)$$

が得られる。この結果は元の f が級数で表されることや微分と無限和の順序交換が可能なことや無限回微分していることから関数 f がある種の性質をみたす必要があるが、指数関数や対数関数や三角関数などの重要な関数でこの性質を満たすことが知られている。よって次の式が得られる。

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + \dots = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}x^k,$$

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots + \frac{1}{n!}x^n + \dots = \sum_{k=0}^{\infty} \frac{1}{k!}x^k,$$

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k} \quad (|x| < 1).$$

これを用いて $x \rightarrow 0$ のとき $e^x \rightarrow 1+x$, $\log(1+x) \rightarrow x$ などの近似式が成立する。問題の状況に応じて必要な次数までとって近似する。なお 1 次のときは平均値の定理を用いて近似式を導くことができる。

6 ガンマ関数とベータ関数

確率分布を考えるときにガンマ関数とベータ関数が出てくることがあり、説明する。

ガンマ関数の定義は、 $a > 0$ に対して

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$$

である。この積分は有限の値に確定することが知られている。なお自然数 n について $\Gamma(n) = (n-1)!$ となることがわかり、自然数以外でもこの関数は定義できるため、階乗の一般化と言われる。

ベータ関数の定義は、 $a, b > 0$ に対して

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

となる関数である。この積分も有限の値に確定する。先ほどのガンマ関数を用いて

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

と表される。