

カントロビッチ不等式と統計学

数理的・計量的な学問において不等式は様々な役割を果たしている。直接評価することが困難な量に対してその近似値や上・下界を与える実用的な不等式もあれば、理論や概念の本質に関わる不等式もある。後者の例としては、統計学における「相関係数」が挙げられる。よく知られる通り、相関係数は2つの変数間の直線的関係の強さの尺度であり、これを r と表せば、 r は $-1 \leq r \leq 1$ なる範囲に値を取り、変数間の直線的関係が強くなるほど1または-1に近づく。そして完全な直線的関係が成り立てば $r = \pm 1$ となる。これらの性質は、よく知られたコーシー・シュワルツ不等式 $(\sum_{i=1}^n a_i b_i)^2 \leq (\sum_{i=1}^n a_i^2)(\sum_{i=1}^n b_i^2)$ から殆ど直ちに導かれるものである。相関係数の概念の殆どはこの不等式によって出来ていると言ってよい。本稿では、古典的不等式の中で、現在でも様々な観点から研究され応用されているカントロビッチ不等式を取り上げ、それが統計学で果たしている役割を平易な例を通して紹介する。

n 個の正の数 a_1, a_2, \dots, a_n が与えられたとする。周知の通り、その算術平均 A と調和平均(=逆数の平均の逆数) H との間には「調和平均 \leq 算術平均」なる関係がある：

$$H = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{a_i} \right)^{-1} \leq \frac{1}{n} \sum_{i=1}^n a_i = A \quad (1)$$

カントロビッチ不等式を一旦この文脈で紹介しておく、(1)式のある種の逆の不等式(すなわち A の H に対する相対的な上界を与える不

等式)であり、 $m \leq a_1, \dots, a_n \leq M$ を満たす任意の正の数 m と M に対して、

$$\frac{1}{n} \sum_{i=1}^n a_i \leq \frac{(m+M)^2}{4mM} \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{a_i} \right)^{-1} \quad (2)$$

と表される(実際はもっと一般的であるが、それについては後述する)。 m と M のシャープな選び方は m を a_1, \dots, a_n の最小値とし、 M を最大値とすることである。係数 $(m+M)^2/4mM$ は、 m と M の算術平均 $(m+M)/2$ と相乗平均 \sqrt{mM} の比の2乗に等しい。あるいは、算術平均 $(m+M)/2$ と調和平均 $\{(m^{-1}+M^{-1})/2\}^{-1} = 2mM/(m+M)$ の比とみることも出来る。また、この係数を m と M の関数とみて $k(m, M)$ と置くと、次のような性質(対称性や不変性)が簡単に示せる。

$$\begin{aligned} k(m, M) &= k(M, m) = k(m^{-1}, M^{-1}), \\ k(m, M) &= k(cm, cM) \text{ for any } c > 0. \end{aligned} \quad (3)$$

さて、同一条件の下で独立に行った n 回の観測から得られた観測値 Y_1, \dots, Y_n があり、各 Y_i は真の平均 μ (未知かつ共通)と誤差 ε_i (観測不能)の和として、

$$Y_i = \mu + \varepsilon_i \quad (i = 1, \dots, n)$$

と書けるものとする。誤差は平均が0で分散 1 が σ^2 (未知かつ共通)であるとする。真の平均 μ は、 n 個の観測値の平均 $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = (Y_1 + \dots + Y_n)/n$ によって推定されることが多い。実際、 \bar{Y} は μ の推定量として様々な長所や最適性を持つ。最も基本

的な長所は、 \bar{Y} に偏りが無い（不偏である）こと、すなわち \bar{Y} の期待値が μ に等しい（ $E(\bar{Y}) = \mu$ ）ことである。実はもっと顕著な結果が成り立ち、 \bar{Y} は線形かつ不偏な推定量の中で最も推定精度が良い、つまり最も分散が小さい。これは一つの最適性であり、最良線形不偏性と呼ばれる。ここで、 μ の推定量 $\hat{\mu}$ が線形であるとは、 $\hat{\mu}$ が Y_1, \dots, Y_n の加重和（線形形式）で書けること、すなわち $\hat{\mu} = c_1 Y_1 + \dots + c_n Y_n = \sum_{i=1}^n c_i Y_i$ という形で書けることである。そして、線形推定量 $\hat{\mu} = \sum_{i=1}^n c_i Y_i$ が不偏であるためには、 $\sum_{i=1}^n c_i = c_1 + \dots + c_n = 1$ となることが必要十分である。線形不偏推定量 $\hat{\mu} = \sum_{i=1}^n c_i Y_i$ の分散は $\sigma^2 \sum_{i=1}^n c_i^2 = \sigma^2 (c_1^2 + \dots + c_n^2)$ となることが知られている。従って最小の分散を持つ推定量を求める問題は、 $\sum_{i=1}^n c_i = 1$ なる条件の下で $\sum_{i=1}^n c_i^2$ を最小にする（ c_1, \dots, c_n ）を求める問題に等しく、これは簡単に解けて、 $c_1 = \dots = c_n = 1/n$ のとき最小、すなわち平均 \bar{Y} が最良線形不偏推定量であることが分かる。その分散は σ^2/n に等しい。

これまで各 $Y_i = \mu + \varepsilon_i$ に含まれる誤差の分散を共通（全て σ^2 ）として議論してきたが、観測時点の天候などといった観測者にコントロールできない要因により、条件が不均一となることがしばしばある。これを表現するため、各 ε_i の分散を σ_i^2 とし（ $i=1, \dots, n$ ）、これらは必ずしも等しくないとする。この場合、観測値によってばらつき具合が異なるため、信頼性の度合いの異なる観測値が混在することとなる。このような場合でも μ の推定に平均 \bar{Y} を用いることは出来る。実際、 \bar{Y} は依然として不偏である。しかし、 \bar{Y} は分散がばらばらの n 個の観測値を等しいウェイト（ $1/n$ ）で扱っている点に不合理

があり、最適な推定量とはならない。最適な推定量を導くため、各 ε_i の分散 σ_i^2 について、それらの比は既知であるとしよう（例えば $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1 : 2 : 1$ ）。より正確に、各 σ_i^2 は未知の $\sigma^2 (>0)$ と既知の $a_1, \dots, a_n (>0)$ によって

$$\sigma_i^2 = \sigma^2 a_i \quad (i=1, \dots, n)$$

と表現出来るものと仮定する。このとき、分散の比は $\sigma_1^2 : \dots : \sigma_n^2 = a_1 : \dots : a_n$ となる。 $a_1 = \dots = a_n = 1$ とすると分散は均一となり、これまでの議論に帰着する。最適性すなわち最良線形不偏性を持つ推定量は、 σ_i^2 の大きさに応じて Y_i のウェイトを変えた

$$\hat{\mu} = \frac{1}{\left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)} \sum_{i=1}^n \frac{1}{\sigma_i^2} Y_i = \frac{1}{\left(\sum_{i=1}^n \frac{1}{a_i}\right)} \sum_{i=1}^n \frac{1}{a_i} Y_i \quad (4)$$

で与えられる。2つの推定量 \bar{Y} と $\hat{\mu}$ の推定精度をそれらの分散で比較すると、 \bar{Y} の分散は $\sigma^2/n, \dots, \sigma_n^2/n$ の算術平均に等しく、 $\hat{\mu}$ の分散は調和平均となる。従って、(1) - (3) 式より

$$\left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)^{-1} \leq \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \leq \frac{(m+M)^2}{4mM} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)^{-1} \quad (5)$$

$\hat{\mu}$ の分散 \leq \bar{Y} の分散 \leq 係数 \times $\hat{\mu}$ の分散

が得られる。ここで、 m と M は $m \leq a_1, \dots, a_n \leq M$ を満たす任意の正の数である。また、 \bar{Y} と $\hat{\mu}$ の分散はそれぞれ

$$\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = \frac{\sigma^2}{n} \times \left(\frac{1}{n} \sum_{i=1}^n a_i\right), \quad \left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)^{-1} = \frac{\sigma^2}{n} \times \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{a_i}\right)^{-1}$$

と分解出来、誤差の分散が均一な場合の推定量の分散（第1因子）と不均一性による影響を表す部分（第2因子）との積に分けることが出来る。

上で $\sigma_1^2, \dots, \sigma_n^2$ の比を既知と仮定したが、

応用上の多くの場合でそれは未知であり、その場合 $\hat{\mu}$ は利用出来ない。未知の比を推定し (4) 式に代入することもあるし、単に平均 \bar{Y} を用いることもある。平均 \bar{Y} を用いたとすると、(5) 式は、比が既知であれば利用出来たであろう最適な推定量 $\hat{\mu}$ に比べ、 \bar{Y} をがどれほど推定精度において劣るか、すなわちどれほど分散が大となるかを示している。例えば、 $1 \leq a_1, \dots, a_n \leq 2$ である場合は、 $m=1, M=2$ として、 $(m+M)^2/4mM=9/8=1.125$ が得られ、平均 \bar{Y} の分散は最適な推定量 $\hat{\mu}$ の分散の1.125倍になり得ることが分かる。

さて、一般にカントロピッチ不等式と言うと次のようなものを指すことが多い²： A を $n \times n$ 正値定符号行列とし、その固有値を $\lambda_1 \leq \dots \leq \lambda_n$ と置く。このとき $x^T x = 1$ なる任意の $n \times 1$ ベクトル x に対して、

$$(x^T A^{-1} x)^{-1} \leq x^T A x \leq \frac{(m+M)^2}{4mM} (x^T A^{-1} x)^{-1} \quad (6)$$

が成り立つ。但し x^T はベクトルや行列の転置を表す。また、 m と M は $m \leq \lambda_1 \leq \lambda_n \leq M$ なる任意の正の数である。あるいは下式の右の不等式

$$1 \leq x^T A x \leq \frac{(m+M)^2}{4mM} (x^T x)^2 \quad (7)$$

のようにも表現される。なお、(7) 式においては $x^T x = 1$ なる条件は不要である。簡単のため、以下では $m = \lambda_1, M = \lambda_n$ とする。

Marshall and Olkin (1990)³ は、カントロピッチ不等式の行列版として、任意のフルランクの $n \times p$ 行列 $X = (x_{ij})$ (但し $p \leq n$) に対して、

$$(X^T A^{-1} X)^{-1} \leq_L (X^T X)^{-1} X^T A X (X^T X)^{-1} \leq_L \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n} (X^T A^{-1} X)^{-1} \quad (8)$$

が成り立つことを示している。ここで、 $A \leq_L B$ は $B - A$ が非負値定符号という意味である。

この式は (5) 式の多次元版である。実際、

(8) 式において $p=1, X = (1, \dots, 1)^T$ とし、 A を対角要素 $\sigma_1^2, \dots, \sigma_n^2$ を持つ対角行列とすると、

(8) 式は (5) 式に帰着する。(8) 式は線形回帰モデル

$$Y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (i=1, \dots, n)$$

において、誤差が必ずしも独立ではなく、また均一分散でもない場合における、回帰係数の推定問題に 응용可能である。その議論は上述の平均の推定問題と共通であるから、ポイントを述べるに止める。誤差 $(\varepsilon_1, \dots, \varepsilon_n)$ の分散共分散行列を A とすると、(8) 式の中辺の行列は最小2乗推定量の分散共分散行列に等しく、最左辺の行列は最良線形不偏推定量の分散共分散行列に等しいことが示せる。従って、上と同様の議論により、最右辺の係数 $k(A) = (\lambda_1 + \lambda_n)^2 / 4\lambda_1 \lambda_n$ は、最小2乗推定量が最良線形不偏推定量に比べてどれほど推定精度が悪くなり得るかを表す量である。係数 $k(A)$ は

$$k(A) = k(A^{-1}), \quad k(A) = k(cA) \text{ for any } c > 0$$

を満たす。これは (3) 式に対応している。

カントロピッチ不等式は様々な観点から研究され、現在でも作用素論や行列解析などの分野で多くの論文が発表されている。統計学の立場からも、これまで行列 A が正値定符号とはならない場合への拡張や、推定精度を1次元の尺度 (行列のトレース、行列式、ノルムなど) で評価した場合の研究など様々なものが発表されてきたが、近年はほぼ結果が出尽くした感があり、ややインパクトが薄いようである。しかし、他分野での発展如何では再び統計

学と接点を持つこともあり得るから、目を離すことは出来ない。

註

- ¹ 分散という概念に馴染みのない方は、「分散は確率変数が平均周りにどれくらいばらつくかを評価する量であり、確率変数の平均が μ 、分散が σ^2 乗ならば、その変数はおよそ95%の確率で $\mu \pm 2\sigma$ の範囲（例：平均10、分散9ならば 10 ± 6 ）に収まる」として解釈されたい。これは正規分布の場合だが、この目安が通用する分布も多い。
- ² 文献としてはR.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press (1985) が定番である。統計学の立場から書かれたものとしてはS. Puntanen, G.P.H. Styan and J. Isotalo, *Matrix Tricks for Linear Statistical Models*, Springer (2011) がある。いちいち挙げないが和文文献（紀要論文や解説記事など）にも優れたものがあり、本稿執筆に際して参考になった。
- ³ A.W. Marshall and I. Olkin, Matrix versions of the Cauchy and Kantorovich inequalities, *Aequationes Mathematicae*, 40 (1990), 89-93.

倉田 博史 (くらた ひろし)

[生年月日] 1967年生

[専門領域] 数理統計学、計量経済学、行列論

[著書・論文] ここ数年は、*Linear Algebra and its Applications*誌、*Journal of Statistical Planning and Inference*誌、*Annals of the Institute of Statistical Mathematics*誌などに論文が掲載されている。詳細は http://www.geocities.jp/h_kurata_statistics/ を参照のこと。

[所属] 総合文化研究科・教養学部、教授

[所属学会] 日本統計学会、日本数学会、*Journal of Japan Statistical Society*誌 Editor-in-Chief、*Annals of the Institute of Statistical Mathematics*誌 Co-Editor