

SSE-DP-2024-3

t 統計量の分布が双峰型となる場合

国友直人・西颯人・薛玉傑

統計数理研究所

2024年7月

SSE-DP(ディスカッションペーパー・シリーズ)は以下のサイトから無料で入手可能です。

<https://stat-expert.ism.ac.jp/training/discussionpaper/>

このディスカッション・ペーパーは、関係者の討論に資するための未定稿の段階にある草稿である。著者の承諾なしに引用・複写することは差し控えられたい。

SSE-DP-2024-3

On Bimodality of t-Statistic

by

Naoto Kunitomo, Hayato Nishi and Xue Yujie

The Institute of Statistical Mathematics

July 2024

(Summary)

Many statisticians use the t-distribution for t-statistic under the assumption of normal population. The distribution of t-statistics can be bimodal when the population distribution is Cauchy. Although there were some studies on this issue by Fiore, Hajivalliliou and Phillips (2010, *Econometric Journal*), and Logan, Mallows, Rice and Shepp (1973, *Annals of Probability*), their analysis are difficult to be understood for applied statisticians. We discuss the mathematical foundations for applications and also report some simulation results when the population is stable distributions.

統計エキスパート演習 (統計基礎を巡る話題 2024):

「t 統計量の分布が双峰型となる場合」¹

2024 年 7 月 26 日

国友直人・西颯人・薛玉傑

要約

統計数理研究所が推進している統計エキスパート養成プログラムの中で slack 上で三輪哲久特任教授が提起した R プログラムによるシミュレーションを巡って、t 統計量の分布が双峰となる場合についての議論をまとめたものである。コーシー分布を含む安定分布についてシミュレーション及び数理的基礎についての検討結果を報告する。

1 はじめに (国友)

統計検定 2 級の教科書「統計学基礎」3 章では多くの教科書と同様に t 統計量についての標準的な説明がある。正規分布にしたがう母集団からの標本平均は母集団の分散が未知の時には標本から計算される分散で基準化すると、その分布は正規分布ではなく t 分布にしたがうことが説明されている。また統計学の歴史を紐解くとダブリン (アイルランド) にあるビール会社ギネス社の技術職員だった William Gosset による t 統計量の分布の導出 (Biometrika, 1905) がその後の近代的な統計学の展開の鍵として (こうした業績の評価は R.A. Fisher から始まると思われるが) しばしば説明されている。

例えば日本の統計家の間で良く引用されている竹村彰通「現代数理統計学」では互いに独立で同一の分布に従う確率変数列 X_1, X_1, \dots, X_n について、もし母集団が正規分布 $N(\mu, \sigma^2)$ であるなら標本平均 $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ について μ を既知としたときの T_n 統計量

$$T_n = \frac{\sqrt{n}[\bar{X}_n - \mu]}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}}} \quad (1.1)$$

が自由度 $n-1$ の t 分布にしたがうことが説明されている。ここで自由度 m の t 分布の密度関数は基準化定数 $c(m)$ とおくと $f(x) = c_m [1 + x^2/m]^{-(m+1)/2}$ ($-\infty < x < \infty$) で与えられる。続いて同書 page 81 では正規母集団でなくても標本不偏分散は

$$V_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1} \quad (1.2)$$

は n が十分に大きいと

$$V_n \xrightarrow{p} \sigma^2 \quad (1.3)$$

¹統計エキスパート slack 上で行われた統計的推測の基礎についての素朴な疑問や議論の中から統計エキスパート養成事業にとり有用と思われる内容をまとめたメモである。三輪哲久特任教授の問題提起 (R プログラムと図 1) より議論が始まったので同教授の許可のもとに原プログラムも掲示しておく。

と確率収束 (convergence in probability) する。したがって、母集団が必ずしも正規分布でなくても標本平均について中心極限定理 (Central Limit Theorem, CLT) より

$$\sqrt{n}[\bar{X}_n - \mu] \xrightarrow{w} N(0, \sigma^2) \quad (1.4)$$

と分布の意味での収束、分布収束あるいは弱収束 (weak convergence) するので、統計量 T_n についても標準正規分布へ収束して

$$T_n \xrightarrow{w} N(0, 1) \quad (1.5)$$

となることが説明されている。こうした統計学における標準的とされている書籍における説明は、(CLT が成り立つ条件などを巡る古典的確率論について内容の正確な理解はさておき) 今日でも統計的方法を利用する応用家の一つのよりどころとなっていることは間違いないだろう。

こうした統計量 T_n の分布についての説明があるにも拘わらず、母集団が正規分布でない場合に t 統計量の分布が歪むこと、例えば二山型になることはあり得ないのだろうか？こうした素朴ではあるが基本的な問題について統計エキスパート養成事業のメンターである三輪哲久教授は短い R プログラムと次のような事実を同事業の slack 上で指摘した。母集団がコーシー分布 Cauchy(0,1) にしたがう場合、シミュレーションにより T_n 統計量の分布は図 1 の様に双峰型になる！(R プログラムは付論 A に掲載した。)

近年ではフリーの統計計算ソフトウェア R、Python などによりこの種のシミュレーションを実際に行うことはかなり容易である。この指摘を受けて様々な実験が可能であるが、分析対象を標準的な i.i.d. 確率変数としても、次のような幾つかの基本的な疑問が生じる。

(i) Cauchy 分布は自由度 1 の t 分布である。母集団分布が仮に自由度 n の t 分布である場合、 T_n 統計量の分布はどうなるだろうか？自由度は必ずしも正整数でない実数でも定義は可能である。

(ii) Cauchy 分布は安定分布族 (stable distribution) の母数 $\alpha = 1$ の場合であることが知られている。(定義は付論を参照のこと。) 安定分布族は $0 < \alpha \leq 2$ のパラメータを持ち、 $\alpha = 2$ が正規分布となることが知られている。母集団分布が仮に安定分布である場合、 T_n 統計量の分布はどうなるだろうか？なお、安定分布族については付論 B でその定義と性質を簡単に述べておくが、統計学で用いられている安定分布にしたがう標準的な乱数発生法については 4 節を参照されたい。

(iii) 統計的データ分析における応用上では T_n 統計量の分布が t 分布や正規分布とはかなり異なる場合、すなわち漸近正規性が成り立たない場合はあり得るのだろうか？その場合にはどの様な統計的方法が有用なのだろうか？

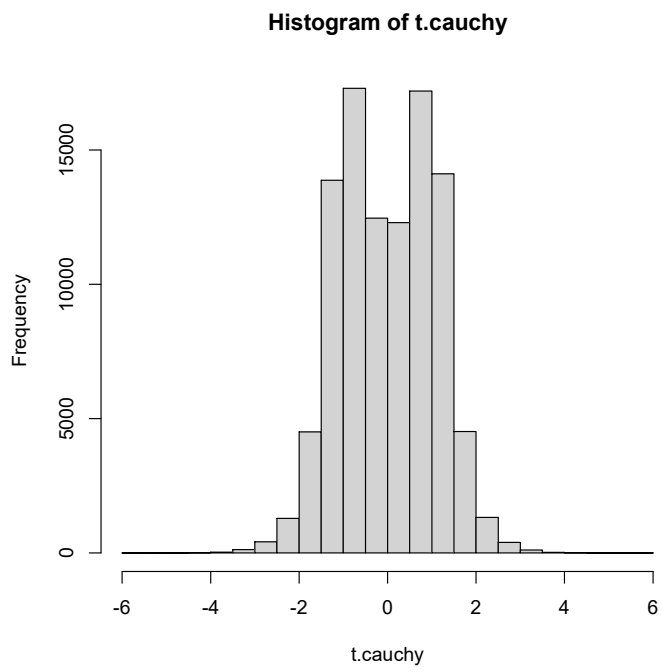


図 1: t 統計量の分布 (シミュレーション)

2 t 統計量の分布が双峰となる実験 (西)

三輪教授が作成した R によるシミュレーション・プログラムを修正、Python プログラムによる幾つかのシミュレーション実験を行った。念のために作成したプログラムは付論 A に掲載して添付しておいた。

2.1. (i) についてのシミュレーション実験 :

t 分布の自由度とサンプルサイズをいろいろ変えて ($n = 10, 30, 100, 1000$) シミュレーションを実行してみると、母集団が $df=1$ (コーシー分布) だけは双峰になり、サンプルサイズを大きくしても双峰性は維持される。すなわちサンプルサイズの影響は小さい。平均値への影響が大きい外れ値の発生確率を考えると、上側と下側で均等に発生する確率よりどちらかに偏る確率の方が高いからこうなるのかもしれないとも考えられる。ここで「平均値への影響が大きい外れ値」をどう定義すればよいのか検討の余地がある。

2.2. (ii) についてのシミュレーション実験

安定分布 (stable distribution) についてもサンプルサイズ ($n = 10, 30, 100, 1000$) として母集団が安定分布となる場合のシミュレーション実験を行ってみた。なお一般の安定分布の場合には分布の歪みも考えられるが、今回は対称分布に限定して実験を行った。母数 $\alpha = 1.5$ 以上だと谷が消え、単峰性があるようである。ただし安定分布の非対称性を示す指標である $\beta = 0$ に固定、ここでは $1 \leq \alpha \leq 1.5$ の結果を示しておいた。この実験における結果 $\alpha = 1.5$ についてはこうした現象について言及している文献は今のところ見つかっていない。更に検討する価値はあるだろう。

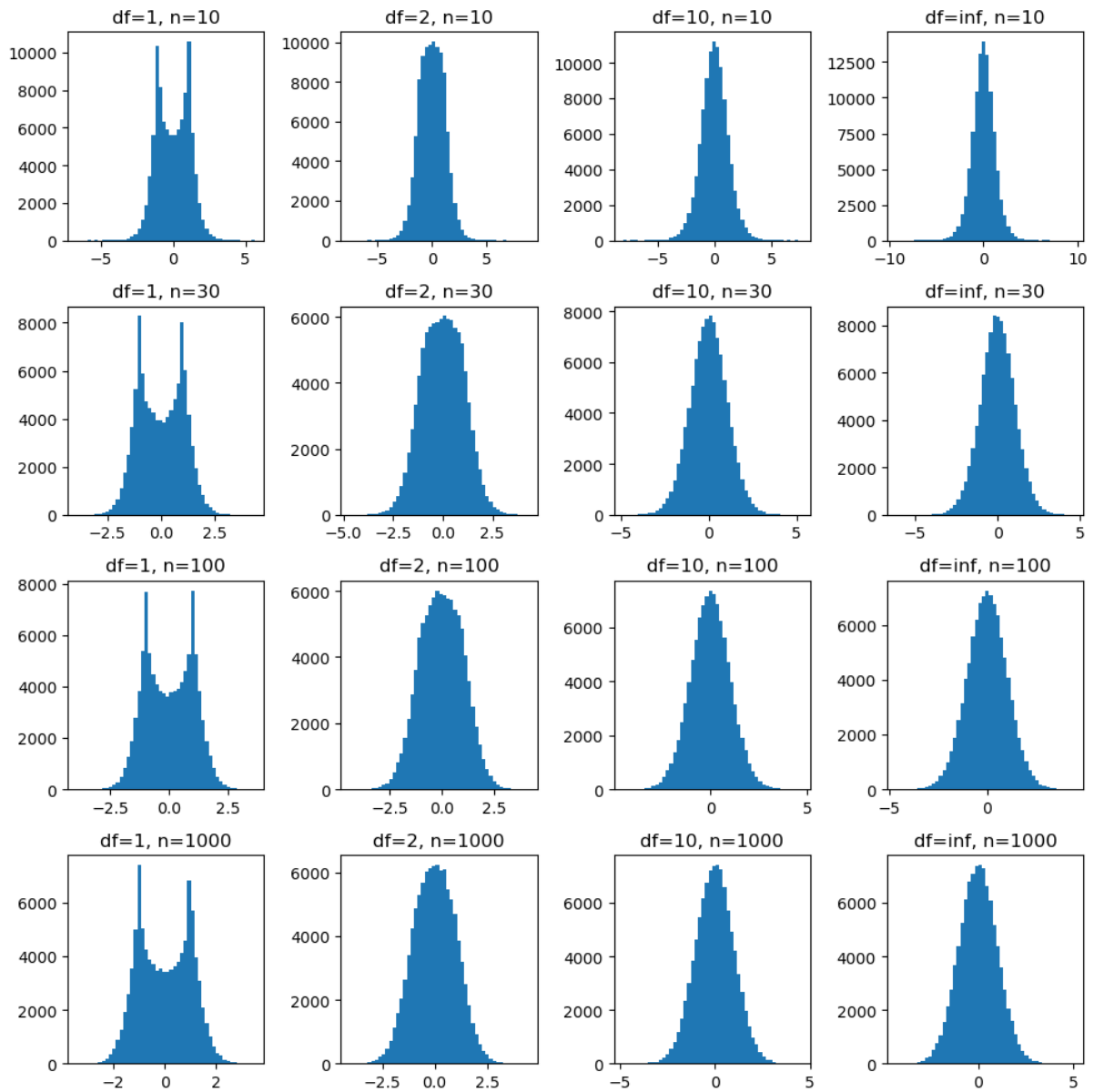


図 2: t 統計量の分布 (シミュレーション)

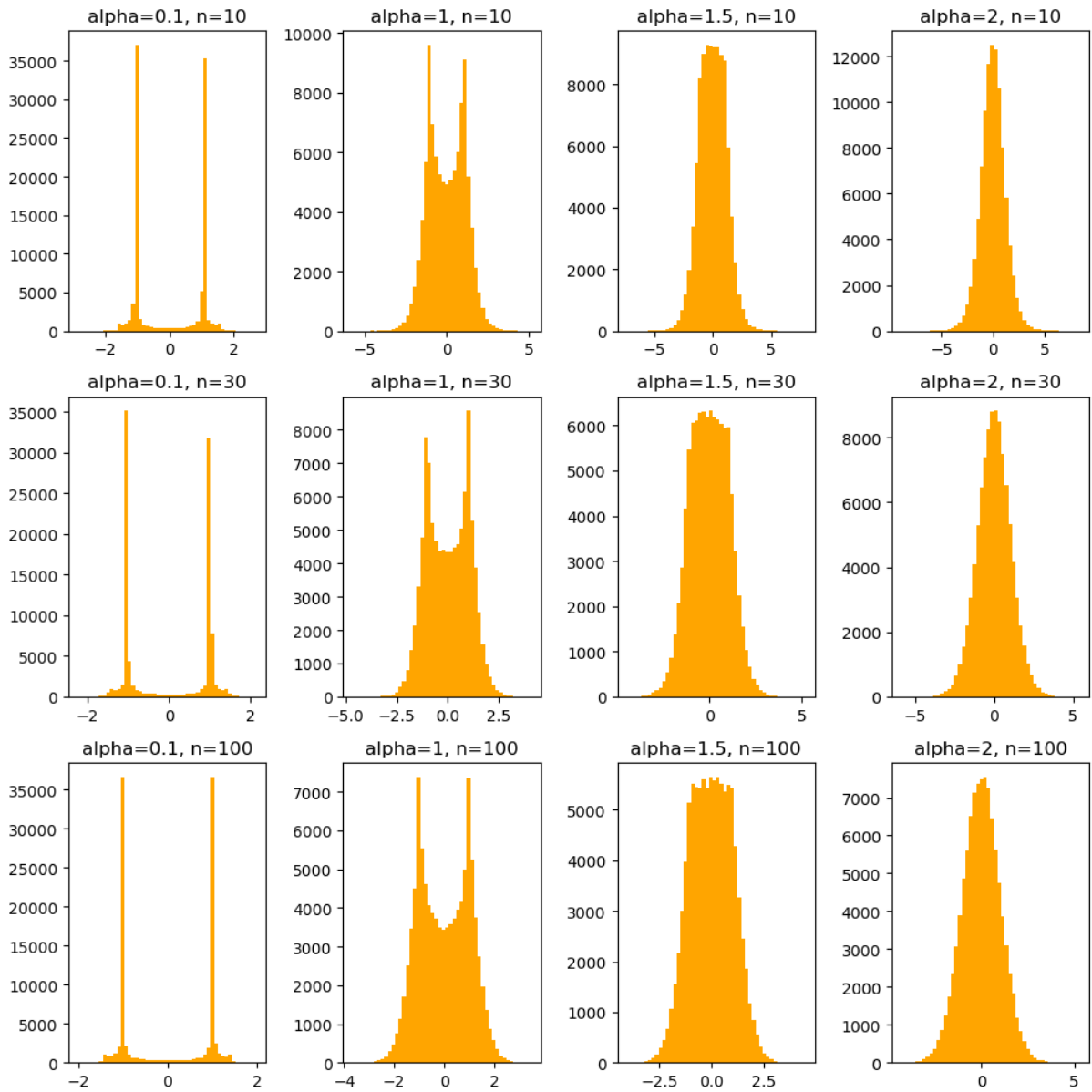


図 3: t 統計量の分布 (シミュレーション)

3 漸近分布の数理について (薛)

i.i.d. 確率変数列が Cauchy 分布にしたがう場合の t 統計量の分布については既に [3] が議論している。この研究はファイナンス分野における裾の厚い分布の統計分析における応用問題の議論から問題を考察したようである。ただし [3] が利用している数理的議論はそのまま直ぐに理解できるとは言い難いので、特殊関数についてよく知られている文献 [2] および確率論の文献 [4] に基づきなるべく分かりやすい解説を試みる。

2階常微分方程式 (second-order ordinary differential equation)

$$xy'' + (c - x)y' - ay = 0 \quad (3.1)$$

の解の形式

$$y = b_1\Phi(a, c; x) + b_2\Psi(a, c; x), \quad (3.2)$$

を考察する。ここで $\Phi(a, c; x) = {}_1F_1(a; c; x)$ は第一種の合流型超幾何関数 (confluent hypergeometric function of first kind) と呼ばれているが、 $\Psi(a, c; x) := \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt} t^{a-1} (1+t)^{c-a-1} dt$ for $\operatorname{Re} a, \operatorname{Re} x > 0$ は第二種の合流型超幾何関数 (confluent hypergeometric function of second kind) と呼ばれている。これらの関数には

$$\Psi(a, c; x) = \frac{\Gamma(1-c)}{\Gamma(a-c+1)} \Phi(a, c; x) + \frac{\Gamma(c-1)}{\Gamma(a)} x^{1-c} \Phi(a-c+1, 2-c; x), \quad (3.3)$$

という関係がある (特殊関数についての文献 [2]. p.257 を参照). なお一般の超幾何関数 (generalized hypergeometric function) とは ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{k=0}^\infty \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{z^k}{k!}$ により定める。ここで記号 $(a)_k = a(a+1) \cdots (a+k-1)$, $a_0 = 1$ により定める。

超幾何関数を利用して Cauchy 分布からの T_n -統計量の性質を調べよう。 X_1, \dots, X_T を Cauchy $(0, 1)$ からの i.i.d. 標本とする。密度関数は $f(x) = \frac{1}{\pi(1+x^2)}$ ($-\infty < x < \infty$) であるが、 $\lim_{x \rightarrow \infty} x^2 f(x) = \lim_{x \rightarrow \infty} x^2 f(-x) = \frac{1}{\pi}$ を満足している。

ここで $U_T = \frac{\sum_{k=1}^T X_k}{T}$, $V_T^2 = \sum_{i=1}^T X_i^2/T^2$ および $S_T = U_T/V_T$ として、統計量 V_T^2 の特性関数を考察する。i.i.d. にしたがう確率変数 X とすると

$$\begin{aligned} cf_{V_T^2}(t) &= \{cf_{X^2/T^2}(t)\}^T \\ &= \{E \exp(i|X|^2 t/T^2)\}^T \\ &= \left\{ \int_{-\infty}^{\infty} \frac{\exp(itx^2/T^2) dx}{\pi(1+x^2)} \right\}^T \\ &= \left\{ \int_0^{\infty} \frac{\exp(itr/T^2) dr}{\pi r^{1/2}(1+r)} \right\}^T \\ &= \{A(t)\}^T \quad (\text{say}) \end{aligned} \quad (3.4)$$

と表現できる。ここで (3.3) より

$$A(t) = \lim_{\varepsilon \rightarrow 0^+} \frac{\Gamma(\frac{1}{2})}{\pi} \Psi\left(\frac{1}{2}, \frac{1}{2}; -it/T^2 + \varepsilon\right), \quad (3.5)$$

であるから

$$A(t) = \frac{\Gamma(\frac{1}{2})}{\pi} [\Gamma(\frac{1}{2})\Phi(\frac{1}{2}, \frac{1}{2}; -it/T^2) + \frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{(-it)^{1/2}}{T} \Phi(1, \frac{3}{2}; -it/T^2)] = 1 + \frac{-2}{\pi^{1/2}} \frac{(-it)^{1/2}}{T} + o(\frac{1}{T}). \quad (3.6)$$

となる。ここで $\Gamma(z) = \int_1^\infty \frac{t^{z-1}}{e^t} dt + \sum_{n=0}^\infty \frac{(-1)^n}{n!} \frac{1}{n+z}$ は拡張されたガンマ関数 Γ であり、通常のガンマ関数 $\Gamma(z)$ の定義域 $\text{Re } z > 0$ から複素平面上 (ゼロおよび負の整数を除く) に連続的 (解析的) に拡張した関数を意味する²。

さらに極限をとると $\lim_{T \rightarrow \infty} cf_{V_T^2}(t) = \exp[\frac{-2}{\pi^{1/2}}(-it)^{1/2}]$ となる。

ここで $-it$ を極座標表示すると $-it = |t| \exp(-i \text{sgn}(t)\pi/2)$ であるから

$$(-it)^{1/2} = |t|^{1/2} \exp(-i \text{sgn}(t)\pi/4) = |t|^{1/2} \cos(\pi/4)(1 - i \text{sgn}(t) \tan(\pi/4)). \quad (3.7)$$

したがって Lévy's 連続性定理 (continuity theorem) より統計量 V_T^2 の極限分布の特性関数は $\exp\{\frac{-2}{\pi^{1/2}}[|t|^{1/2} \cos(\pi/4)(1 - i \text{sgn}(t) \tan(\pi/4))]\}$ で与えられる。この特性関数の形から $\alpha = 1/2$ の安定分布であることが分かる。

次に (U_T, V_T^2) の特性関数を考察する。まず

$$cf_{U_T, V_T^2}(s, t) = \{cf_{X/T, X^2/T^2}(s, t)\}^T. \quad (3.8)$$

$$cf_{X/T, X^2/T^2}(s, t) = \int_{-\infty}^\infty \frac{e^{isx/T + itx^2/T^2}}{\pi(1+x^2)} dx. \quad (3.9)$$

となることを利用する。ここで変換 $x \rightarrow (r, h)$ (ただし $r = x^2, h = \text{sgn}(x)$) を利用する。このとき

$$cf_{X/T, X^2/T^2}(s, t) = \int_0^\infty \frac{\exp(itr^2/T^2)}{\pi(1+x^2)} \int_h \frac{1}{2} \exp(is\sqrt{r}h/T) dh dr. \quad (3.10)$$

となる。ここで Taylor 展開を e^z に施すと次の結果が得られる。

$$\frac{1}{2} \int_h e^{(is/T)\sqrt{r}h} dh = \sum_{k=0}^\infty \frac{((is/T)^2/4)^k r^k}{k!(\frac{1}{2})_k}.$$

したがって

$$\begin{aligned} cf_{X/T, X^2/T^2}(s, t) &= \int_0^\infty \frac{\exp(itr^2/T^2)}{\pi(1+x^2)} \sum_{k=0}^\infty \frac{((is/T)^2/4)^k r^k}{k!(\frac{1}{2})_k} dr \\ &= \frac{1}{\pi} \sum_{k=0}^\infty \frac{((is/T)^2/4)^k \Gamma(k + \frac{1}{2})}{k!(\frac{1}{2})_k} \Psi(k + \frac{1}{2}, k + \frac{1}{2}, -it/T^2) \end{aligned} \quad (3.11)$$

となる。(3.3) を用いて $\Psi(k + \frac{1}{2}, k + \frac{1}{2}, -it/T^2)$ に $\Gamma(\frac{1}{2} - k)\Phi(k + \frac{1}{2}, k + \frac{1}{2}; -it/T^2) + \frac{\Gamma(k - \frac{1}{2})}{\Gamma(\frac{1}{2} + k)}(-it/T^2)^{1/2-k}\Phi(1, \frac{3}{2} - k; -it/T^2)$, $\Gamma(\frac{1}{2} - k) = \frac{\pi}{(-1)^k \Gamma(k + \frac{1}{2})}$, $\Phi(k + \frac{1}{2}, k + \frac{1}{2}; -it/T^2) =$

²ここでは Erdélyi [2], Chapter 1 の表現を採用した。正実数に対するガンマ関数を複素関数へ拡張するときには幾つかの等価な定義が存在する。複素解析についての標準的文献としては Ahlfors, L.V. (1979), *Complex Analysis*, 3rd Edition, McGraw-Hill, Inc., 高橋礼司 (1989), 「複素解析」, 東京大学出版会、読みやすい書籍としては志賀浩二 (1989), 「複素数 30 講」, 朝倉書店 などがある。

e^{-it/T^2} を代入すると

$$cf_{X/T, X^2/T^2}(s, t) = \sum_{k=0}^{\infty} \frac{(-is/T)^2/4)^k}{k!(\frac{1}{2})_k} e^{-it/T^2} + \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{((is/T)^2/4)^k \Gamma(k - \frac{1}{2})}{k!(\frac{1}{2})_k} (-it/T^2)^{1/2-k} \Phi(1, \frac{3}{2} - k; -it/T^2) \quad (3.12)$$

となる。ここで $\Gamma(k - \frac{1}{2}) = \Gamma(-\frac{1}{2})(-\frac{1}{2})_k$ となることに注意すると,

$$\begin{aligned} cf_{X/T, X^2/T^2}(s, t) &= 1 + o(1/T) + \frac{\Gamma(-\frac{1}{2})}{\pi} \sum_{k=0}^{\infty} \frac{((is/T)^2/4)^k (-\frac{1}{2})_k}{k!(\frac{1}{2})_k} (-it/T^2)^{1/2-k} (1 + o(1/T)) \\ &= 1 + \frac{\Gamma(-\frac{1}{2})}{\pi} \left(\sum_{k=0}^{\infty} \frac{((s^2/4it)/4)^k (-\frac{1}{2})_k}{k!(\frac{1}{2})_k} \right) (-it/T^2)^{1/2} + o(1/T) \\ &= 1 + \frac{\Gamma(-\frac{1}{2})}{\pi} \Phi(-\frac{1}{2}, \frac{1}{2}; s^2/4it) (-it/T^2)^{1/2} + o(1/T) \quad (3.13) \end{aligned}$$

が得られる。したがって極限では

$$\lim_{T \rightarrow \infty} cf_{U_T, V_T^2}(s, t) = \exp\left[\frac{-2}{\pi^{1/2}} \Phi(-\frac{1}{2}, \frac{1}{2}; s^2/4it) (-it)^{1/2}\right]. \quad (3.14)$$

となるので、二次元確率変数 (U_T, V_T^2) の極限分布の特性関数が得られた。そこで次に $S = U/V$ の極限分布を導こう。

再び特性関数

$$E \exp(iUs + iV^2t) = \exp\left[\frac{-2}{\pi^{1/2}} \Phi(-\frac{1}{2}, \frac{1}{2}; s^2/4it) (-it)^{1/2}\right] = \exp[B(s, t)] \quad (\text{say}). \quad (3.15)$$

を考察する。ここで正実数 a, b をとり $t = is^2a$ および $t = is^2b$ とおき、 $0 < s < \infty$ で積分すると

$$\int_0^{\infty} E \frac{1}{s} e^{iUs} \{e^{-V^2s^2a} - e^{-V^2s^2b}\} ds = \int_0^{\infty} \frac{1}{s} (e^{sB(1, ia)} - e^{sB(1, ib)}) ds \quad (3.16)$$

となる。さらに $Vs = t$ とおくと

$$\begin{aligned} \int_0^{\infty} E \frac{1}{t} e^{i(U/V)t} \{e^{-t^2a} - e^{-t^2b}\} dt &= \int_0^{\infty} \frac{1}{s} (e^{sB(1, ia)} - e^{sB(1, ib)}) ds \\ \int_0^{\infty} \frac{1}{t} \phi(t) \{e^{-t^2a} - e^{-t^2b}\} dt &= \int_0^{\infty} \frac{1}{s} (e^{sB(1, ia)} - e^{sB(1, ib)}) ds, \quad (3.17) \end{aligned}$$

となる。ここで $\phi(t)$ は S の特性関数である。(3.17) の両辺を a で微分すると、

$$\int_0^{\infty} t \phi(t) e^{-t^2a} dt = -B'(1, ia) \int_0^{\infty} e^{sB(1, ia)} ds = \frac{B'(1, ia)}{B(1, ia)}. \quad (3.18)$$

さらに $(\pi a)^{-1/2} e^{-s^2/4a}$ を乗じて、 a について $0-\infty$ で積分し、恒等式

$$\frac{t}{\pi^{1/2}} \int_0^{\infty} \exp[-s^2/4a - at^2] a^{-1/2} da = \exp(-st), \quad s > 0, t > 0, \quad (3.19)$$

を利用すると $s > 0$ に対して

$$\int_0^{\infty} \phi(t)e^{-st} dt = \int_0^{\infty} (\pi a)^{-1/2} e^{-s^2/4a} \frac{B'(1, ia)}{B(1, ia)} da. \quad (3.20)$$

が得られる。

確率変数 S の密度関数を $g(y)$ とする。すなわち、特性関数の逆変換 (実部)

$$g(y) = \lim_{s \rightarrow iy} \operatorname{Re} \frac{1}{\pi} \int_0^{\infty} \phi(t)e^{-st} dt \quad (3.21)$$

で与えられる。ここで [3] の (4.35) を利用して $\int_0^{\infty} (\pi a)^{-1/2} e^{-s^2/4a} \frac{B'(1, ia)}{B(1, ia)} da$ を $s = iy$ として連続的に変形すると、積分値は

$$g(y) = \lim_{s \rightarrow iy} \operatorname{Re} \frac{1}{\pi} \int_0^{\infty} \phi(t)e^{-st} dt = \operatorname{Re} \int_0^{\infty/i^{1/2}} e^{y^2 t^2/2} \tilde{D}(t) dt \quad (3.22)$$

と表現できる³。ただし $\tilde{D}(t) = (\frac{2}{\pi^3})^{1/2} \{1 + te^{t^2/2} \int_0^t e^{-u^2/2} du\}^{-1}$ で与えられる。

さらに変換 $t \rightarrow re^{-i\pi/4}$ を用いると

$$\int_0^{\infty/i^{1/2}} e^{y^2 t^2/2} \tilde{D}(t) dt = \int_0^{\infty} e^{-y^2 r^2 i/2} \tilde{D}(re^{-i\pi/4}) e^{-i\pi/4} dr \quad (3.23)$$

と表現できる。

ここで [4] の (iii) より ($y \rightarrow 1$ のとき), $g(y) \sim \frac{1}{\pi^2} \log \frac{1}{|1-y|}$ となる。また Logan et al. [4] の補題 A, 補題 B を利用すると $y \rightarrow \infty$ のとき $g(y) \sim (\frac{2}{\pi})^{1/2} \tau_0 e^{-1/2 y^2 \tau_0^2}$ となる。ただし $\tilde{D}(-i\tau_0) = \infty$ である。

最後に t 統計量の漸近分布を考察しよう。 t 統計量 $t_T = \frac{U_T}{(V_T^2 - \frac{1}{T} U_T^2)^{1/2}}$ とする。 $U_T = \bar{X} \sim Cauchy(0, 1)$ のとき $V_T^2 - U_T^2/T = V_T^2 - O_p(1/T)$ である。したがって $t_T = S_T + O_p(1/T)$ となる。すなわち t 統計量の極限分布は S_T の極限である S の分布に一致することが分かる。

注 1 Logan et al. [4] の (4.35) について :

Logan et al. [4] より $B(1, ia)$ は $\mu = 0$ としたときの $\tilde{\psi}(a)$ (Logan et al. [4] の (4.41))、すなわち

$$B(1, ia) = \int_{-\infty}^{\infty} [e^{-x^2 a} e^{ix} - 1] \frac{1}{x^2} dx \quad (3.24)$$

である。したがって

$$B'(1, ia) = \int_{-\infty}^{\infty} -e^{-x^2 a} e^{ix} dx. \quad (3.25)$$

となる。ここで

$$a = 1/(2b^2), \quad x = \xi b \quad (3.26)$$

³この変形は Cauchy の積分定理による。Logan et al. [4] に極の位置に関する正当化が説明されているがかなり技術的議論なのでここでは省略する。なお (3.22) の右辺は複素変数 t を極座標 (r, θ) で表示すると、 $i^{1/2} = [\exp(i\pi/2)]^{1/2} = \exp(i\pi/4)$ より $t = r * \exp(-i\pi/4)$ での実数 r の範囲は $(0, \infty)$ 、偏角 $\theta = -\pi/4$ に設定という意味である。ここで $dt = \exp(-i\pi/4) dr$ と表現できる。

を代入すると右辺 (3.20) は

$$\begin{aligned} \int_0^\infty (\pi a)^{-1/2} e^{-s^2/4a} \frac{B'(1, ia)}{B(1, ia)} da &= \int_0^\infty \sqrt{\frac{2}{\pi}} e^{-s^2 b^2/2} \frac{\int_{-\infty}^\infty -e^{-\xi^2/2} e^{i\xi b} d\xi}{\int_{-\infty}^\infty [e^{-\xi^2/2} e^{i\xi b} - 1] \frac{1}{\xi^2} d\xi} db \\ &= \pi \int_0^\infty e^{-s^2 b^2/2} \tilde{\mathcal{D}}(b) db, \end{aligned} \quad (3.27)$$

となるが、ただし $\tilde{\mathcal{D}}(b) = \sqrt{\frac{2}{\pi^3}} \frac{\int_{-\infty}^\infty e^{-\xi^2/2} e^{i\xi b} d\xi}{\int_{-\infty}^\infty [1 - e^{-\xi^2/2} e^{i\xi b}] \frac{1}{\xi^2} d\xi}$ である。 $\sqrt{\frac{1}{2\pi}}$ を乗じた分子 $\int_{-\infty}^\infty e^{-\xi^2/2} e^{i\xi b} d\xi$ は正規分布の特性関数

$$\int_{-\infty}^\infty e^{-\xi^2/2} e^{i\xi b} d\xi = \sqrt{2\pi} e^{-b^2/2} \quad (3.28)$$

である。分母は

$$\begin{aligned} \int_{-\infty}^\infty [1 - e^{-\xi^2/2} e^{i\xi b}] \frac{1}{\xi^2} d\xi &= \int_{-\infty}^\infty [1 - e^{-\xi^2/2} \cos(\xi b)] \frac{1}{\xi^2} d\xi \\ &= -[1 - e^{-\xi^2/2} \cos(\xi b)] \frac{1}{\xi} \Big|_{-\infty}^\infty + \int_{-\infty}^\infty \frac{be^{-\xi^2/2} \sin(\xi b)}{\xi} d\xi + \int_{-\infty}^\infty e^{-\xi^2/2} \cos(\xi b) d\xi \\ &= 0 + b \int_{-\infty}^\infty e^{-\xi^2/2} \left(\int_0^b \cos \xi u du \right) d\xi + \sqrt{2\pi} e^{-b^2/2} \\ &= b \int_0^b \sqrt{2\pi} e^{-u^2/2} du + \sqrt{2\pi} e^{-b^2/2} \end{aligned} \quad (3.29)$$

である。したがって $\tilde{\mathcal{D}}(t) = \left(\frac{2}{\pi^3}\right)^{1/2} \{1 + te^{t^2/2} \int_0^t e^{-u^2/2} du\}^{-1}$ が得られる。

注2： 複素積分による評価について (極と留数の定理など)

Logan et al. [4] 補題 A より負の虚軸 t 上で関数 $e^{y^2 t^2/2} \tilde{\mathcal{D}}(t)$ は $\tilde{\mathcal{D}}(t) = \infty$ となるただ一つの極 (pole) t_0 が存在する。そこで閉曲線 $C(\rho)$ 、曲線 $C_1(\rho)$ (0 から $\rho \exp -i\pi/2$ および第 4 象限の t_0 の周りを修正), $C_2(\rho)$ ($\rho \exp -i\pi/2$ から $\rho \exp -i\pi/4$), C_3 ($\rho \exp -i\pi/4$ から 0) をとる。このとき

$$\int_{C(\rho)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt = \int_{C_1(\rho)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt + \int_{C_2(\rho)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt + \int_{C_3(\rho)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt \quad (3.30)$$

となる。Logan et al. [4] 補題 B から $\rho_n^* \rightarrow \infty$ を $\int_{C_2(\rho_n^*)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt$ ((3.30) の右辺) がゼロに収束して、 $C_2(\rho_n^*)$ と $C_2(\rho_{n+1}^*)$ の間に唯一の極があるようにとれる。この時の極と経路 $C(\rho)$ を図 4 で表す。

ここで留数の定理 (The residue theorem) を用いると

$$\int_{C(\rho_n^*)} \exp(y^2 t^2/2) \tilde{\mathcal{D}}(t) dt = 2\pi i \sum_{i=1}^n \text{Res}(\exp(y^2 t^2/2) \tilde{\mathcal{D}}(t), t_i) \quad (3.31)$$

となるがここで t_i は $C_2(\rho_i^*)$ と $C_2(\rho_{i-1}^*)$ の間にある極である。したがって $n \rightarrow \infty$ のとき (3.30) より

$$\lim_{n \rightarrow \infty} \int_{C(\rho_n^*)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt = \lim_{n \rightarrow \infty} \int_{C_1(\rho_n^*)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt + \lim_{n \rightarrow \infty} \int_{C_3(\rho_n^*)} e^{y^2 t^2/2} \tilde{\mathcal{D}}(t) dt. \quad (3.32)$$

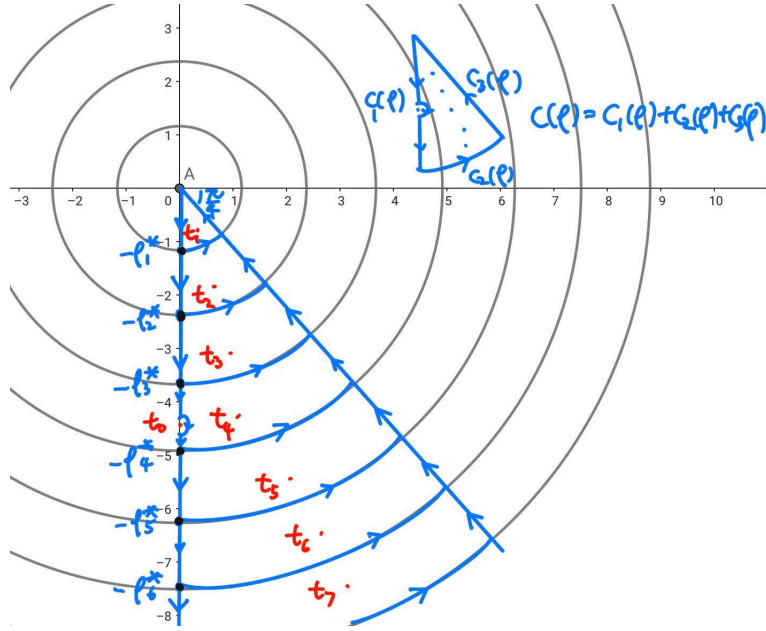


図 4: 極 (poles) と $C(\rho)$

となる⁴。さらに (3.31), (3.32), (3.22) より $C_2(\rho_n^*)$ 上の積分をゼロとして経路の符号を考慮すると、

$$\begin{aligned}
 g(y) &= \text{Real}\left\{-\lim_{n \rightarrow \infty} \int_{C_3(\rho_n^*)} e^{y^2 t^2/2} \tilde{D}(t) dt\right\} \\
 &= \text{Re}\left\{\lim_{n \rightarrow \infty} \int_{C_1(\rho_n^*)} e^{y^2 t^2/2} \tilde{D}(t) dt - 2\pi i \sum_{i=1}^{\infty} \text{Res}(e^{y^2 t^2/2} \tilde{D}(t), t_i)\right\}.
 \end{aligned} \tag{3.33}$$

が得られる。虚軸上で関数 $e^{y^2 t^2/2} \tilde{D}(t)$ の実部は 0、さらに $\lim_{n \rightarrow \infty} \int_{C_1(\rho_n^*)} e^{y^2 t^2/2} \tilde{D}(t) dt$ の実部 $-1/2 \text{Re}\{2\pi i \text{Res}(e^{y^2 t^2/2} \tilde{D}(t), t_0)\}$ は t_0 周辺で修正した経路上での積分値の実部に対応する。Logan et al. [4] の (5.6) より $\text{Res}(e^{y^2 t^2/2} \tilde{D}(t), t_i) = -\sqrt{\frac{2}{\pi^3}} e^{y^2 t_i^2/2} t_i$ for $i = 0, 1, 2, \dots$ となるので⁵

$$g(y) = \text{Re}\left\{2\pi i \sum_{i=1}^{\infty} \sqrt{\frac{2}{\pi^3}} e^{y^2 t_i^2/2} t_i\right\} + \text{Re}\left\{\pi i \sqrt{\frac{2}{\pi^3}} e^{y^2 t_0^2/2} t_0\right\} \tag{3.34}$$

となる。Logan et al. [4] 809 項の議論より $|y| > 1$ のとき $g(y)$ は有界、さらに数値的検討により Logan et al. [4] は $y \rightarrow \pm\infty$ のとき $g(y) \sim (\frac{2}{\pi})^{1/2} \tau_0 e^{-1/2 y^2 \tau_0^2}$ となると推測している。なお τ_0 は $\tau_0 > 0$, $\tilde{D}(-i\tau_0) = \infty$ (虚軸上の極) である。また $|y| < 1$ では Logan et al. [4](4.36) の議論 (実数 t をとると $\tilde{D}(t) = O(\exp(-t^2/2))$ など) より

$$g(y) = \int_0^{\infty} e^{y^2 t^2/2} \tilde{D}(t) dt, \tag{3.35}$$

$t \rightarrow \infty$ のとき

$$e^{y^2 t^2/2} \tilde{D}(t) \sim \frac{1}{\pi^2} e^{(y^2-1)t^2/2} t^{-1} \tag{3.36}$$

⁴ここで有限個の極が存在するとき可積分であることが示せるので Logan et al. [4] が議論しているように極限をとることが正当化できると思われるので詳しい議論は省略する。

⁵ここで孤立特異点 t_i ($i = 1, \dots, n$) における留数 (residue) は項 $e^{y^2 t_i^2/2} (2/\pi^3)^{1/2} [-t_i/(t - t_i)]$ の係数 $e^{y^2 t_i^2/2} (2/\pi^3)^{1/2} (-t_i)$ を利用した。

と近似できる。したがって

$$\begin{aligned}
g(y) &\sim \int_{1-y}^{\infty} \frac{1}{\pi^2} e^{(y^2-1)t^2/2} t^{-1} dt + C \\
&\sim \frac{1}{2\pi^2} \int_{(1-y)^2}^{\infty} e^{(y^2-1)s/2} / s ds + C \\
&\sim \frac{1}{2\pi^2} e^{(y^2-1)s/2} \ln s \Big|_{(1-y)^2}^{\infty} - \int_{(1-y)^2}^{\infty} (y^2-1)/2e^{(y^2-1)s/2} \ln s ds + C \quad (3.37)
\end{aligned}$$

と近似できる。ただし $\int_0^{\infty} e^{-s} \ln s ds = -\gamma$ となることに注意すると (ここで γ は Euler の定数) 近似的に

$$g(y) \sim -\frac{1}{\pi^2} \log|1-y| \quad (3.38)$$

($y \rightarrow 1$ のとき) となる。同様に $y \rightarrow -1$ のとき

$$g(y) \sim -\frac{1}{\pi^2} \log|1+y| \quad (3.39)$$

となることが分かる。

4 幾つかの注意点 (国友)

本稿では統計学ではよく知られた T_n 統計量の標本分布についてあらためて3名が検討した結果を報告したが、この間に気が付いた統計エキスパートにとり有用と考えられる幾つかの注意点をまとめて述べておく。

(i) 安定分布にしたがう確率変数の生成

正規分布にしたがう確率変数列を生成する方法としては Box-Muller 法が統計学では良く知られている。一様分布 $U(0, 1)$ にしたがう互いに独立な確率変数 U_1, U_2 より変換 $X_1 = \mu + \sigma \sqrt{-2 \log(U_1)} \cos 2\pi U_2$, $X_2 = \mu + \sigma \sqrt{-2 \log(U_1)} \sin 2\pi U_2$, は独立な $N(\mu, \sigma^2)$ にしたがう二つの正規確率変数となる。

安定分布にしたがう確率変数列については Chamber-Mallosw-Stuck 法 (Chamber et al. (1976)) が統計学では標準的である。 Θ を一様分布 $U(-\pi/2, \pi/2)$ 、 W を $\lambda = 1$ の指数分布に独立にしたがうとする。

1. $0 < \alpha \leq 2, \alpha \neq 1$ のとき

$$Z = \frac{\sin \alpha \Theta}{(\cos \Theta)^{1/\alpha}} \left[\frac{\cos(\alpha-1)\Theta}{W} \right]^{(1-\alpha)/\alpha}, \quad (4.1)$$

$\alpha = 1$ のとき $Z = \tan \Theta$ とする。このとき $Z \sim S(\alpha, 0, 1)$, 安定分布 $S(\alpha, 0, 1)$ にしたがう。

2. 一般の (歪みのある) 安定分布を含む場合は、 $-1 \leq \beta \leq 1$, $\theta_0 = \arctan(\beta \tan(\pi\alpha/2))/\alpha$ ($\alpha \neq 1$) とおく。 $0 < \alpha \leq 2, \alpha \neq 1$ のとき

$$Z = \frac{\sin \alpha(\theta_0 + \Theta)}{(\cos \alpha \theta_0 \cos \Theta)^{(1/\alpha)}} \left[\frac{\cos(\alpha\theta_0 + (\alpha-1)\Theta)}{W} \right]^{(1-\alpha)/\alpha}, \quad (4.2)$$

$\alpha = 1$ のとき

$$Z = \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta \Theta \right) \tan \Theta - \beta \log \left(\frac{\frac{\pi}{2} W \cos \Theta}{\frac{\pi}{2} + \beta \Theta} \right) \right] \quad (4.3)$$

とする。このとき $Z \sim S(\alpha, \beta, 1)$, 安定分布 $S(\alpha, \beta, 1)$ にしたがう。(証明は Nolan (2020)3 章に与えられている。)

(ii) T_n 統計量の分布

竹村彰通「現代数理統計学」など日本の統計学関係者の間では定評がある教科書における中心極限定理 (CLT) の説明は十分とは言えない。古典的な確率論では大数の法則や CLT について詳しく調べられているが、独立な確率変数列についての CLT では Lindeberg 条件が基本的であり、よく知られた十分条件として確率変数が二次積率 $\mathbf{E}[X^2] < \infty$ を持つことが要請される。コーシー分布の密度関数は $f(x) = (1/\pi)(1/(1+x^2))$ ($-\infty < x < \infty$) であるから期待値 $\mathbf{E}[|X|]$ は発散する。自由度 n の t 分布は $n = 1$ の場合はコーシー分布、 $n = 2$ の場合は期待値は有限であるが二次積率は発散する。安定分布としては $\alpha = 2$ は正規分布となるが、 $\alpha = 1$ はコーシー分布に対応している。したがって $1 < \alpha < 2$ の場合については幾つかの興味深い問題が生じ得る。CLT が成立するとは限らないが、再生性 (和の分布が元の分布に属するという性質) があることから古典的な確率論ではかなり研究されている⁶。しかし、コーシー分布からの統計量の分布についてはあまり注目されてこなかったようである。

このレポートの 3 節において解説しているように数理的には実はかなり難しい議論が必要となることもこの問題が十分に理解されていない原因と考えられる。第 1 節で述べたように T_n の漸近正規性が成り立つためには (i) 分子について CLT が成り立つ, (ii) 分母が確率収束する、ことが必要となる。コーシー分布の場合には位置母数 $\mu = 0$ としておくと (i) コーシー分布からの標本平均 $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ は再びコーシー分布にしたがう。(付論 B を参照。) (ii) $V_n = \sum_{i=1}^n X_i^2/n^2$ は $n \rightarrow \infty$ のときある確率変数 $V (> 0)$ に収束する。この (ii) を示すことは自明ではないので正確に示すことが第 3 節の最初の課題である。分布を明示的に表現できないので特性関数による表現を利用するが、特性関数と分布関数は 1 対 1 に対応する (Levy の反転公式による) ので、特性関数の極限が存在すれば極限分布が存在することが分かる。コーシー分布の場合には V は $\alpha = 1/2$ の安定分布となることが示される。次に \bar{X}_n と V_n の同時分布の極限が存在することが 3 節の第 2 の内容である。こうして得られた同時特性関数から統計量 T_n の分布を導くには有限個とは限らない極 (poles) に留数の定理 (residue theorem) を適用する必要があるようである。したがってこの辺からなぜ問題が簡単でないことも理解できる。さらに最終的な目的である T_n 統計量の極限分布は存在はするものの、CLT による正規分布とは様相がかなり異なることが分かる。例えば極限分布の密度関数はある点において発散しうる、などの論点の説明が 3 節の内容である。

こうした Cauchy 分布を巡る結果は既に Fiero et al. (2010) の Theorem 2.1 として述べられているが、彼らの証明は不完全と思われ (あるいは分かりにくい) ので、本稿ではその簡単な導出を試みている。ところで、問題を検討する過程で実はこうした結果は Logan et al. (1973) により既に安定分布に関する様々な結果の一部として既に得られていたことが判明した。しかしながら、Logan et al. (1973) の論文は (普通の統計家には) かなり高度な複素解析・特殊関数の議論を多用して数理的結果を導くことに加えて、数学関係者を対象とする確率論の専門学術誌の論文であり、普通の応用統計家にはかなり近寄りがない

⁶CLT や安定分布などを含め古典的確率論においてよく知られている内容を説明している教科書の例としては Durrett, R. (1991), "Probability: Theory and Example," Duxbury を挙げておく。

感がある。そこで3節の後半では Logan et al. (1973) による結果を可能な範囲でより分かりやすく理解できるように彼らの論文には説明がない幾つかの議論と解釈を展開した。Logan et al. (1973) の図5では $\alpha = 1.5$ の場合について密度関数は発散しないが僅かに双峰となる場合があるようである。また極限分布を通常の変換公式を利用した積分計算により数値計算を試みると、しばしば数値が発散した。これは極限分布の密度関数が場所により発散しうることに符合すると解釈できるかもしれない。

本稿での考察から結論として得られた T_n 統計量の漸近分布が双峰型になり得る、という論点は少なくともこれまで多くの日本の統計家の間で知られていたとは言い切れない。応用上の可能性を含めて一つの重要な結果と言えるだろう。

(iii) 安定分布と応用

安定分布の株価への応用可能性を 1963 年⁷ に指摘したのは物理学者の Mandelbrot であり、その後、一時期は E. Fama に代表されるファイナンス分野の研究者が注目した。その後、こうした流行はかなり下火になったとはいえ、再び金融リスク管理の問題においても確率分布の裾 (稀に起きる事象 rare events) の評価に関連して注目されている。株価と云ったファイナンス分野に限らず工学や理学分野などでの統計データ分析における応用可能性については例えば Nolan(2020) が積極的な立場から展望を行っている。

(iv) 幾つかの残された課題

正規分布やコーシー分布を含む確率分布として安定分布族の重要性は理解できたとしても、その先はあり得るだろうか？確率変数列の和の極限分布としては正規分布やコーシー分布が挙げられるが、確率論では安定分布族を含む、和についての全ての可能な極限分布を表現する無限分解可能分布 (付論を参照されたい) があり、安定分布は一つの特例な場合であることが知られている。こうした一般化により統計的データ分析における新たな手がかりが得られる可能性は皆無ではないと思われる。

文献

- [1] Chamber, J., C. Mallows and B. Stuck, (1976). *A method for simulating stable random variables*, *Journal of the American Statistical Association*, Vol. 71 (354), 340-344. (Correction 82(2), 497-503).
- [2] Erdélyi, A. (1953). *Higher Transcendental Functions*, Vol.1. New York: McGraw-Hill.
- [3] Fioero, C., Hajivalliliou V. A., & Phillips P.C.B. (2010), Bimodal t-ratios, *Econometric Journal*, 13, (2), 271-289.
- [4] Logan, B. F., Mallows, C. L., Rice, S. O., & Shepp, L. A. (1973). Limit distributions of self-normalized sums. *The Annals of Probability*, 1(5), 788-809.
- [5] Nolan, J. (2020). *Univariate Stable Distributions: Models for Heavy Tailed Data*, Springer.
- [6] 統計学基礎 (改訂版), 日本統計学会編, 2015, 東京図書.

⁷Mandelbrot, B. (1963), "The variation of some speculative prices," *Journal of Business*, 26,394-419.

[7] 現代数理統計学 , 竹村彰通, 2020, 学術図書出版.

付論 A : シミュレーション・プログラム

```
# By Miwa 2024-4-30
#
#set.seed(1)
n <- 10
N <- 100000
t.cauchy <- numeric(N)
for(i in 1:N) {
  x <- rcauchy(n)
  t.cauchy[i] <- sqrt(n)*mean(x)/sqrt(var(x))
}
#ヒストグラム.
hist(t.cauchy)

#set.seed(1)
n <- 30
N <- 100000
t.rt <- numeric(N)
for(i in 1:N) {
  x <- rt(n,2)
  t.rt[i] <- sqrt(n)*mean(x)/sqrt(var(x))
}
#ヒストグラム.
hist(t.rt)
```

```

# Nishi 2024-5-7
#
%%
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

# %%
def gen_t_stats(n, rng, df=1, n_sim=100000):
    if df == np.infty:
        x = rng.standard_normal(size=(n, n_sim))
    else:
        x = rng.standard_t(df=df, size=(n, n_sim))
    t = np.sqrt(n) * np.mean(x, axis=0) / np.sqrt(np.var(x, axis=0))
    return t

rng = np.random.default_rng(123)

dfs = [1, 2, 10, np.inf]
ns = [10, 30, 100, 1000]

fig, axes = plt.subplots(len(ns), len(dfs), figsize=(10, 10))

for i, n in enumerate(ns):
    for j, df in enumerate(dfs):
        axes[i, j].set_title(f"{df=}, {n=}")
        axes[i, j].hist(gen_t_stats(n, rng, df), bins=50)
fig.tight_layout()
fig.show()

# %%

rng = np.random.default_rng(123)

dfs = [1.2, 1.4, 1.6, 1.8]
ns = [10, 30, 100]

fig, axes = plt.subplots(len(ns), len(dfs), figsize=(10, 10))

for i, n in enumerate(ns):
    for j, df in enumerate(dfs):
        axes[i, j].set_title(f"{df=}, {n=}")
        axes[i, j].hist(gen_t_stats(n, rng, df), bins=50)
fig.tight_layout()
fig.show()

# %%
def gen_t_stats_stable(n, rng, alpha=1, n_sim=100000):
    dist = stats.levy_stable(alpha=alpha, beta=0.0)
    x = dist.rvs(size=(n, n_sim), random_state=rng)
    t = np.sqrt(n) * np.mean(x, axis=0) / np.sqrt(np.var(x, axis=0))
    return t

rng = np.random.default_rng(123)

alphas = [0.1, 1, 1.5, 2]
ns = [10, 30, 100]

fig, axes = plt.subplots(len(ns), len(alphas), figsize=(10, 10))

```

```
for i, n in enumerate(ns):
    for j, alpha in enumerate(alphas):
        axes[i, j].set_title(f"{alpha=}, {n=}")
        axes[i, j].hist(gen_t_stats_stable(n, rng, alpha), bins=50, color="orange")
fig.tight_layout()
fig.show()

# %%
```

```

# Nishi 2024-5-7
# %%
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

# %%
def gen_t_stats(n, rng, df=1, n_sim=100000):
    if df == np.infty:
        x = rng.standard_normal(size=(n, n_sim))
    else:
        x = rng.standard_t(df=df, size=(n, n_sim))
    t = np.sqrt(n) * np.mean(x, axis=0) / np.sqrt(np.var(x, axis=0))
    return t

rng = np.random.default_rng(123)

dfs = [1, 2, 10, np.inf]
ns = [10, 30, 100, 1000]

fig, axes = plt.subplots(len(ns), len(dfs), figsize=(10, 10))

for i, n in enumerate(ns):
    for j, df in enumerate(dfs):
        axes[i, j].set_title(f"{df=}, {n=}")
        axes[i, j].hist(gen_t_stats(n, rng, df), bins=50)
fig.tight_layout()
fig.show()

# %%

rng = np.random.default_rng(123)

dfs = [1.2, 1.4, 1.6, 1.8]
ns = [10, 30, 100]

fig, axes = plt.subplots(len(ns), len(dfs), figsize=(10, 10))

for i, n in enumerate(ns):
    for j, df in enumerate(dfs):
        axes[i, j].set_title(f"{df=}, {n=}")
        axes[i, j].hist(gen_t_stats(n, rng, df), bins=50)
fig.tight_layout()
fig.show()

# %%
def gen_t_stats_stable(n, rng, alpha=1, n_sim=100000):
    dist = stats.levy_stable(alpha=alpha, beta=0.0)
    x = dist.rvs(size=(n, n_sim), random_state=rng)
    t = np.sqrt(n) * np.mean(x, axis=0) / np.sqrt(np.var(x, axis=0))
    return t

rng = np.random.default_rng(123)

alphas = [0.1, 1, 1.5, 2]
ns = [10, 30, 100]

fig, axes = plt.subplots(len(ns), len(alphas), figsize=(10, 10))

for i, n in enumerate(ns):

```

```
    for j, alpha in enumerate(alphas):
        axes[i, j].set_title(f"{alpha=}, {n=}")
        axes[i, j].hist(gen_t_stats_stable(n, rng, alpha), bins=50, color="orange")
fig.tight_layout()
fig.show()

# %%
```

付論 B : 安定分布について⁸

2項分布の近似に端を発した中心極限定理 (CLT) は歴史的には様々な方向に拡張が試みられた。しかし、例えば確率変数の分散 $\mathbf{E}[Z^2] = \infty$ のときは和の分布はどうなるのであろうか？例えばコーシー分布をとると $\mathbf{E}[|Z|] = \infty$ で与えられる。この場合には特性関数⁹は $\varphi(t) = \mathbf{E}[\exp(itZ)] = e^{-|t|}$ となる (なお $i^2 = -1$ である)。したがって、もし独立な確率変数 Z_i ($i = 1, \dots, n$) がコーシー分布にしたがっていれば和 $S_n = \sum_{i=1}^n Z_i$ の特性関数は $\varphi_{S_n}(t) = e^{-n|t|}$ であるから確率変数 $Y_n = S_n/n$ の特性関数は $\varphi_{Y_n}(t) = e^{-|t|}$ となり再びコーシ分布にしたがう。ここで注目すべき点は基準化は S_n/\sqrt{n} ではなく $S_n/n^{1/\alpha}$ ($\alpha = 1$) となることである。これに対して中心極限定理 (central limit theorems, CLT) は $\alpha = 2$ の場合に対応する。

コーシー分布は期待値も発散する極端な確率分布と考えられるのでもう少し現実的に期待値は定義できるが分散 $\mathbf{E}[Z^2] = \infty$ となる場合を考察しよう。

例 A.1 : 確率分布として $P(Z > z) = P(Z < -z)$ ($z > 0$) および $P(|Z| > z) = z^{-\alpha}$ ($z \geq 1, 0 < \alpha < 2$) を取りあげる。特性関数 $\varphi(t)$ を考察すると

$$\begin{aligned} 1 - \varphi(t) &= \int_1^{\infty} (1 - e^{itz}) \frac{\alpha}{2} z^{-\alpha-1} dz + \int_{-\infty}^{-1} (1 - e^{itz}) \frac{\alpha}{2} |z|^{-\alpha-1} dz \\ &= \alpha \int_1^{\infty} \frac{1 - \cos(tz)}{z^{\alpha+1}} dz \\ &= t^\alpha \alpha \int_t^{\infty} \frac{1 - \cos(u)}{u^{\alpha+1}} du \end{aligned}$$

となる。ここで $\int u^{-\alpha-1} du$ は可積分なので定数 $C = \alpha \int_0^{\infty} (1 - \cos(u))/u^{\alpha+1} du$ とすると $t \rightarrow 0$ のとき $1 - \varphi(t) \sim C|t|^\alpha$ と近似できる。したがって、確率変数 Z_i ($i = 1, \dots, n$) が互いに独立にこの分布にしたがっているとき、和 $S_n = \sum_{i=1}^n Z_i$ の特性関数は

$$\begin{aligned} \mathbf{E}[\exp(it \frac{S_n}{n^{1/\alpha}})] &= \left[\varphi\left(\frac{t}{n^{1/\alpha}}\right) \right]^n \\ &= \left[1 - (1 - \varphi\left(\frac{t}{n^{1/\alpha}}\right)) \right]^n \\ &\rightarrow e^{-C|t|^\alpha} \quad (n \rightarrow \infty) \end{aligned}$$

となる。こうした評価より分布の裾が厚い場合には一般に確率変数 and の極限分布の特性関数は定数 C_2, C_3 を利用して

$$\varphi_Z(t) = \exp \left[C_2 \int_1^{\infty} (e^{itz} - 1) \frac{dz}{z^{\alpha+1}} + C_3 \int_{-\infty}^{-1} (e^{itz} - 1) \frac{dz}{|z|^{\alpha+1}} \right]$$

となることが期待される。

⁸ 「統計エキスパート演習 2023」第 2 章付論の再掲。

<https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2024/01/SSE-DP-2024-1-3.pdf>

⁹ 特性関数は実数 t に対する複素関数 $\varphi(t) = \mathbf{E}[\exp(itZ)]$ で与えられ、分布関数と 1-1 対応する。コーシー分布の特性関数を求めるには複素積分を利用する必要があるが、詳しくは確率論の教科書を参照されたい。

定義 A.1 : 確率変数 Z が無限分解可能分布 (infinitely divisible distribution) にしたがうとは、和の分布が $\mathcal{L}[Z_{n1} + \dots + Z_{nn}] = \mathcal{L}[Z]$ となる互いに独立・同一分布にしたがう確率変数列 Z_{ni} ($i = 1, \dots, n$) が存在することである。

原点での漸近挙動を考慮した表現は次のようにまとめることができる。

定理 A.1 : 無限分解可能分布にしたがう確率変数 Z の特性関数は $C_0, C_1 (> 0)$ を定数として

$$\varphi_Z(t) = \exp \left[itC_0 - \frac{C_1}{2}t^2 + \int (e^{itz} - 1 - itz1_D)\nu(dz) \right] \quad (\text{A.4})$$

と表現される。ここで $D = \{|z| \leq 1\}$, $\int (z^2 \wedge 1)\nu(dz) < \infty, \nu(\{0\}) = 0$ である。 $(\nu(\cdot)$ は Levy 測度と呼ばれている。)

注意 : Levy(レビー) 測度は異なる形で表現されることがある。無限分解可能分布はジャンプ確率過程と密接に関係する¹⁰。一般に無限分解可能分布は重要な確率分布を含んでいる。例で利用した確率分布の特性関数は安定分布 (stable distribution) の特性関数の形となるが、例では中心部分を見捨てたことに注意する必要がある。分散が有限でない場合には分布の中心部分の評価と裾の評価を別々に行い、最終的に組み合わせる必要が生じる。安定分布は次のよう定義され、特性関数で表現されることが知られているが詳細は省略する。

定義 A.2 : 独立で同一分布にしたがう確率変数列 Z_i (i, \dots, n) が確率変数 Z の分布と同一であり、ある定数 a_n, b_n が存在し和の分布 $\mathcal{L}[Z_1 + \dots + Z_n] = \mathcal{L}[a_n Z + b_n]$ となるとき Z の分布は安定分布 (stable distribution) と呼ばれる。

定理 A.2 : (i) 安定分布にしたがう確率変数 Z の特性関数は $C_1, C_2, C_3, 0 < \alpha \leq 2$ を定数として次の形で与えられる。

$$\begin{aligned} \varphi_Z(t) = & \exp \left[itC_0 - \frac{C_1 t^2}{2} + C_2 \int_0^\infty (e^{itz} - 1 - itz1_D) \frac{dz}{z^{\alpha+1}} \right. \\ & \left. + C_3 \int_{-\infty}^0 (e^{itz} - 1 - itz1_D) \frac{dz}{|z|^{\alpha+1}} \right] \end{aligned} \quad (\text{A.5})$$

ただし $D = \{|z| \leq 1\}$ である。

(ii) 特性関数は C_0, d ($d > 0$), θ ($-1 \leq \theta \leq 1$) を定数として次の形になる。

(i) $0 < \alpha < 1, 1 < \alpha \leq 2$ のとき

$$\varphi_Z(t) = \exp \left[itC_0 - d|t|^\alpha \left(1 + i\theta \frac{t}{|t|} \tan \frac{\pi}{2} \alpha \right) \right], \quad (\text{A.6})$$

(ii) $\alpha = 1$ のとき

$$\varphi_Z(t) = \exp \left[itC_0 - d|t| \left(1 + i\theta \frac{t}{|t|} \frac{2}{\pi} \log |t| \right) \right]. \quad (\text{A.7})$$

¹⁰例えば佐藤健一「加法過程」(岩波書店)が詳しい。