統計数理研究所 75 周年記念事業（2019）

極値分布の吸引領域と離散性

Maximum domain of attraction and Discreteness

数理・推論研究系 志村 隆彰 （Takaaki Shimura）

キーワード：極値理論, 極値分布, 確率分布の裾, 最大値吸引領域, 離散化.

近年，東日本大震災のような巨大地震や西日本豪雨のような豪雨など，極端に大きな被害をもたらす自然災害（激甚災害）が続き，大きな社会問題になっている．このような災害をランダムな現象とみなせば，“起こる頻度（確率）は少ないが，一旦起こったときの被害（量）が極端に大きい”ことが特徴である．このような“稀に起こる極端な事象とその起こり方”を研究するのが極値理論 (Extreme Value Theory) である．極値理論では，極端な事象の起こり方を表す確率分布 F の裾（確率）$\bar{F}(x) \equiv 1 - F(x)$ （x より大きい値が起こる確率を表す）の挙動が重要で，x が F の上限 $x_F \equiv \sup\{x : F(x) < 1\}$ に近づくときの 0 への収束が，正規分布のように収束が速いとき裾が軽い，コーシー分布のように遅いとき裾が重いという（heavy tail は最近広く知られるようになってきている）．

確率統計でもっとも重要な定理は中心極限定理であるが，極値理論でこれに当たるのが多数の中の最大値の振る舞いを表す次の定理である．

定理０． $X_1, X_2, \ldots$ を共通の確率分布 F に従う実数値独立確率変数列とし，X_n までの最大値を $M_n \equiv \max\{X_1, \ldots, X_n\}$ とする．このとき，共通分布 F が適当な条件を満たせば，非退化分布 G が存在して，適当な定数列 $a_n > 0$ と $b_n \in \mathbf{R}$ により，

$$\mathcal{L}\left(\frac{M_n - b_n}{a_n}\right) \to G \quad (n \to \infty)$$

となる ($\mathcal{L}()$ はカッコ内の確率変数の分布を，$\to$ は分布の収束を表す)．このときの極限分布 G は極値分布と呼ばれ，３種類ある（フレシェ分布，グンベル分布，（極値）ワイブル分布）．

各極値分布に対して，上記の収束が成り立つような確率変数列の共通分布 F の集合をその極値分布の（最大値）吸引領域といい，分布の裾 $\bar{F}(x)$ が分布の上限に近づくときの漸近挙動で特徴づけられる．本稿では，F として上限が無限の分布を扱うので，極値分布はフレシェ分布あるいはグンベル分布になる．裾の重さに注目すると，パレート分布のような裾の挙動がべきオーダーで重い（正確には裾が正則変動する）ものはフレシェ分布の吸引領域に，指数分布などの裾がより軽いものはグンベル分布の吸引領域に属する．また，分布の連続性の観点からは，正規分布，指数分布，コーシー分布など様々な裾挙動を持つ連続分布がいずれかの極値分布の吸引領域に属する半面，幾何分布やポアソン分布などの裾が重くない離散分布はどの極値分布の吸引領域にもはいらない（指数分布と幾何分布は連続分布か離散分布かという違いであることに注意）．この事実は，吸引領域への属性と離散性との相性が悪いことを示唆している．分布が吸引領域に入らないということは，理論的に，それに従う独立同分布列の最大値の挙動がとらえきれないだけでなく，応用上も，実データが必然的に離散であるから，統計解析の精度に影響することを意味している．

この背景の元，最大値吸引領域と分布の離散性の関係を考察した結果を述べる．実数上の分

布に対して，区間 $(n-1,n]$ の確率を一点 $\{n\}$ に集中させて離散分布（整数値分布）を作る操作を分布の”離散化”と呼び，離散分布に対して，連続分布でその離散化が元の離散分布と一致するものを離散分布の”連続化”と呼ぶ（連続化は一意ではない）．幾何分布は指数分布の離散化，幾何分布の連続化のひとつが指数分布である．このような場合，指数分布は離散化で吸引領域への属性を失い，幾何分布は連続化で属性が回復可能ということにする．

問題は，第一に，吸引領域に属する分布で，離散化によりその属性を保つか，失うかの判別条件を与えること．第二に，連続化で属性が回復可能な分布の特徴付けである．

定理 A. 吸引領域に属する分布が，離散化により，吸引領域に留まるための必要十分条件は，長い裾を持つことである：$\lim_{x\to\infty}\bar{F}(x+1)/\bar{F}(x)=1$ (long-tailed distribution).

定理 B. 分布 F が吸引領域に回復可能であるための必要十分条件は，そのフォン・ミーゼス関数 $\bar{F}_0$ ($\bar{F}$ を滑らかにしたもの）が次を満たすことである．

$$\lim_{n\to\infty}\left(\log\frac{\bar{F}_0(n+1)}{\bar{F}_0(n+2)}\right)^{-1}-\left(\log\frac{\bar{F}_0(n)}{\bar{F}_0(n+1)}\right)^{-1}=0.$$

定理 A から，パレート分布や対数正規分布は離散化で属性を失わないこと，定理 B から幾何分布の他に，ポアソン分布も回復可能であることがわかる（定理 B の吸引領域は実はグンベル分布の吸引領域）．

次に，幾何的分布と呼ばれる離散分布に対して，吸引領域に属する連続化を与える具体的方法を与える．離散分布 F_1 が幾何的分布であるとは，ある正の指数 γ が存在して，$\lim_{n\to\infty}\bar{F}_1(n+1)/\bar{F}_1(n)=e^{-\gamma}$ となるときをいう．また，F が指数的分布であるとは，ある正の指数 γ が存在して，$\lim_{x\to\infty}\bar{F}(x+k)/\bar{F}(x)=e^{-\gamma k}$ となるときをいい，その全体を $\mathcal{L}(\gamma)$ であらわす．$\mathcal{L}(\gamma)$ はグンベル分布の吸引領域に含まれることが知られている．

定理 C. F_1 を指数 $\gamma(>0)$ の幾何的分布，G を $[0,1]$ 上の分布とする．このとき，$F_1*G\in\mathcal{L}(\gamma)$ となるための必要十分条件は，

$$G(x)=\frac{1-e^{-\gamma x}}{1-e^{-\gamma}}\quad(0\le x\le 1)$$

である．

幾何的分布，指数的分布はそれぞれ幾何分布，指数分布の一般化であり，ガンマ分布などが指数的分布である．この定理は，幾何的分布は指数が共通である限り，同じ $[0,1]$ 上の分布を（確率変数の意味で独立に）足すことで，吸引領域へ回復可能であることを主張している．たとえば，幾何分布に従う実データは定理 0 の極限定理は成り立たないが，データにランダムな補正をすることで成り立つようにできるのである．

参 考 文 献

Shimura, T. (2012). Discretization of distributions in the maximum domain of attraction, Extremes **15**, 299-317.

Shimura, T. (2012). Limit distribution of a roundoff error, Statistics and Probability Letters, **82**, 713-719.

高橋 倫也・志村 隆彰. (2016). 「極値統計学」(ISM シリーズ 進化する統計数理５), 近代科学社.