希少事象の数学的研究

志村 隆彰　統計基盤数理研究系 准教授

1　希少事象の研究

日々発生する様々なことは、想定内のことと想定外のことに大別され、後者は、ランダムな現象とみなして考察されることが多い。ランダムに発生するとみなされる事象のうち、発生する確率が小さいものを、希少事象(Rare event)をいう。希少事象はめったに起こらないから、事後対応で十分などと考えがちかもしれないが、この考えが正しくないことは、大地震のような甚大災害（希少事象の典型）を思い浮かべれば直ちにわかる。災害対策の第一は防災であり、防災は、どのような規模の事象が、どれくらいの確率で発生するかを予測し、その予測に基づいて、様々な具体的な対策をとるものであるから、最初にすべきは、希少事象とその発生確率を考えることである。一方で、関心の対象が希少事象という限られたものであるから、地震の例でも、ある程度大きい震度の地震だけが考察の対象となり、震度 1 の地震のような数多くの頻繁に起こる事象は無視されることが多い。こうした事情から、希少事象においては、通常の統計解析で基本的な平均や分散は重要でなく、そもそもそうした量が存在しない、大数の法則や中心極限定理が成立しない状況を扱うことが少なくない。

2　数学的表現と設定

ランダムな現象は、数学的に以下のように表現される。まず、事象は数値とする。地震でいえば、マグネチュードや震度、大雨であれば、雨量が代表的な考察対象になる。次に、事象が発生する（起こる）確率を考える。ランダムな事象を確率変数 X で、その発生確率を 確率分布 F(確率分布関数$F(x) = Pr(X \le x)$(X が $x(\in \mathbf{R})$以下をとる確率) を表す。

ここまでは、ランダムな現象の数学的研究（確率論或いは数理統計学）に共通であるが、対象が希少事象である場合は、x が大きいときの確率分布関数の裾（確率）(tail probability)

$$\bar{F}(x) := 1 - F(x) = Pr(X > x)$$

が希少事象とその確率（起こり方）を数学的に表現するものになる。確率分布F の上限 $x_F := \sup\{t : F(t) < 1\}$ とすると考察対象の現象により、$x_F = \infty$ と $x_F < \infty$ となる場合があり、いずれも上限付近の裾の値が希少事象の発生法則を表現している。xが上限 x_F に近くなれば、全ての分布の裾確率は、0 に近づくが($\lim_{x \to x_F} \bar{F}(x) = 0$) 、希少事象の性質は、裾確率が0に収束（近づく）する速さに反映され、考察は、その速さによって分けて行われる。具体的な確率分布の（0への）収束の速さは、下記のように正規分布では、指数オーダーより速く、指数分布では、指数オーダー、パレート分布ではべきのオーダー（指数オーダーより遅い）といった具合である。

【例】代表的な確率分布の裾の漸近挙動

順に（標準）正規分布、指数分布（パラメータ１)、パレート分布の裾確率。どれも0へ収束するが、収束速度は大きく異なる。

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{1}{2}t^2\right) dt << \exp(-cx),\ c > 0 \quad (x \to \infty),$$

$$\int_x^\infty \exp(-t)\, dt = \exp(-x),$$

$$x^{-\alpha}, \quad \alpha > 0.$$

裾の0への収束が速いことは、稀な現象が起こりにくいことを、遅いことは稀な現象が起こりやすいことを意味し、速いとき、裾が軽い(light tail)、遅いとき裾が重い(heavy tail)という。

コーシー分布は、パレート分布と同じく、べきオーダーの裾をもち、大数の法則や中心極限定理が成り立たない、異端の分布のように見なされることがあるが、応用面で需要の高いのは、極端に大きい事象が高頻度で起こる裾が重い分布に従う現象であり、現実にもそのような分布に従う現象が少なくないことが知られている。

3　極値理論

稀事象を数学的に扱う分野の代表が極値理論(Extreme value theory)である。数多くのデータを表す確率変数列$\{X_1, X_2, \ldots\}$を考え、基本的な設定として、この確率変数列を（共通の分布をFとする）独立同分布列（independent identically distributed (i.i.d)）とする。最初の関心事は、n 項までの（部分）最大値 $M_n := \max\{X_1, X_2, \ldots, X_n\}$ のn が大きいときの挙動である。X_n が n番目の地震のマグネチュードやn日目の降雨量を表すものとすれば、M_n は n までに起こる地震の中で最大のマグネチュードやn日目までの日最大降雨量を意味する。この例でもわかるよに、M_nは災害と強くむつびつく量と見なすことができる。

M_n の挙動について、M_n が$n \to \infty$ のときに分布Fの上限値x_Fに近づくことは直感的に明らかであろう（$\lim_{n\to\infty} M_n = x_F (a.e.)$)。より意味のある$M_n$ の挙動は、独立和 $S_n := \sum_{k=1}^n X_n$ の挙動の記述である大数の法則や中心極限定理で行われるのと同様に、位置とスケールを適当に変換する。つなわち、適当な定数列 $a_n > 0$と$b_n \in \mathbf{R}$ によって、M_nを正規化した $\frac{M_n - b_n}{a_n}$ の挙動を考える。

中心極限定理が独立和S_nの平均を0、分散を1に正規化して、$n \to \infty$ としたとき、正規化された和の分布が標準正規分布に収束するように、正規化された最大値M_nは極値分布と呼ばれる分布に収束することが知られている。

4　離散（整数値）確率分布に対する極値理論

離散確率分布に対する極値理論に関心がある。　実は、上述の極限定理(正規化された最大値M_n の極値分布への収束)分布は、分布 F がポアソン分布や幾何分布のような代表的な離散分布に対しては成り立たない(全ての離散分布というわけではない)。つまり、分布の離散性と既存の極値理論は必ずしも相性がよくないといえる。この傾向は、裾が軽い分布ほど顕著になるが、離散化で漸近挙動が変わらない、吸引領域への属性が保たれるパレート分布のような裾が重い分布であっても推定量の漸近正規性がなくなって、統計解析に支障をきたすことがある。離散分布に対する極値理論と関連する統計解析の改善が目下の研究課題である。

5　共同研究集会

統計数理研究所では、共同利用事業の一環として、共同研究集会を開催しています。うち、当該研究と関係するふたつを紹介します。

【極値理論の工学への応用】７月から８月に２日間開催予定。
【無限分解可能過程に関連する諸問題】
　　１０月から１２月に３日間開催予定。

研究成果報告（共同研究リポート）、過去の開催情報などは、以下のサイトをはじめ、統数研ホームページのイベント欄、及び各種メーリングリストで入手できます。共同研究集会に参加費はなく、どなたでもご自由にできますので、ご関心のある方はご覧ください。

sites.google.com/view/takaakishimura

参考文献

[1] Discretization of distributions in the maximum domain of attraction, Extremes, 15 (2012) 299-317.

[2] Limit distribution of a roundoff error, Statistics and Probability Letters, 82 (2012), 713-719.