

確率分布の裾挙動に関する研究

志村 隆彰 数理・推論研究系 准教授

【極端事象】

ランダムな現象に伴う数値において、値が極めて大きいあるいは小さい事象を極端事象という。現実問題として、極端事象は大変珍しい事象なので、ないものとして扱ってかまわない場合が少なくない。しかし、それとは逆に極端事象こそが意味を持つ場合も多くある。強い台風、大雨に伴う洪水や河川の氾濫、巨大地震など近年多発する甚大災害はその典型で、その対策の重要性は言うまでもないが、数多くの雨や地震のうち、災害を引き起すのは、ごく少数の極端に大規模なものだけである。このようにランダムな現象全体というよりもそのごく一部のみが重要となる状況を考えたい。

【確率分布の裾】

ランダムな数値をあらわす確率変数 X の分布 (関数) $F(x) = Pr(X \leq x)$ に対し、

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

を分布 F の裾 (tail) という。裾は x の関数として、 x 以上の事象が起こる確率を表し、以降、上限が無限の場合 (すべての x に対して、 $\bar{F}(x) > 0$) を考える。 $\bar{F}(x)$ は $x \rightarrow \infty$ のとき 0 に近づくが、この収束の速さが速いとき、裾が軽い (light tail)、遅いとき、裾が重い (heavy tail) という。確率密度関数を見ただけでは同じように見える正規分布とコーシー分布だが、裾の挙動 ($x \rightarrow \infty$ のときの 0 への収束の様子) は全く違い、正規分布の裾は速く、コーシー分布の裾はゆっくりと 0 にいく。正規分布のような裾の軽い分布は、極端に大きい、或いは小さい値をとる確率が少ないため、その分布に従う独立確率変数列に対して、大数の法則や中心極限定理が成り立つ。一方で、コーシー分布のような裾が重く、極端な値を取りやすい分布に対してはそのような定理は成り立たない。また、裾の重い分布と軽い分布では同じような大きさの値をとってもその値に違いが出る。標準正規分布表をよくみれば気が付くが、正規分布に従う乱数のうち、大きいもの、たとえば 2 以上或は 3 以上のものの小数点一桁目 (さらには 2 桁目以降の数字) 以下の数字は 0 や 1 のような小さい数ほど多く、8 や 9 のような大きい数は少ない。この傾向は乱数の数値が大きくなるに従って強くなり、やがて小数点以下の数字、さらには 2 桁目以降の数字のほとんどは 0 になる。一方、重い裾をもつ分布ではこのようなことは起きず、コーシー分布では、数値が大きくなると小数点以下にはどの桁にも 0, 1, ..., 9 の数字が等しく現れるようになる。数値の頭から二桁目以降の数字の現れ方についても、正則変動する裾を持つ分布に対して次の定理が成り立つ。

定理 1 ([1]) 分布 F が指数 $-\alpha (< 0)$ の正則変動をする裾を持つ、すなわち、任意の $\lambda > 0$ に対して、

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(\lambda x)}{\bar{F}(x)} = \lambda^{-\alpha}$$

を満たすと仮定する。 X を分布 F に従う確率変数とし、 N を X の桁数、 K を X の先頭の数字とし、 $Y = 10^{-N+1}X - K$ とおく (X の先頭の数字を除き、それ以外を $[0, 1)$ に写したもの)。このとき、 $0 \leq x \leq 1$ に対し、

$$\lim_{n \rightarrow \infty} P(Y \leq x | K = k, N = n) = \frac{1 - (1 + \frac{x}{k})^{-\alpha}}{1 - (1 + \frac{1}{k})^{-\alpha}} \quad k = 1, 2, \dots, 9.$$

コーシー分布は $\alpha = 1$ の場合に当たり、コーシー分布のような正則変動する裾を持つ分布に従う確率変数の極端に大きな実現値は、その先頭の数字 (0, 1, ..., 9) によって、二桁目以降の数字の分布が異なる特定の分布に収束する。

後半では前半とは別の無限分解可能分布の裾の研究を紹介する。

【無限分解可能分布とそのレヴィ測度の遠方での関係】
 μ が無限分解可能分布とは、任意の自然数 n に対して、分布 μ_n で

$$\hat{\mu}(z) = (\hat{\mu}_n(z))^n$$

となるものがとれるときをいう。正規分布、コーシー分布、複合ポアソン分布など多くの分布が無限分解可能である。ここで、 $\hat{\mu}(z)$ は μ の特性関数 (フーリエ変換) である。正の台をもつものに限れば、その特性関数は、次のようにかける。

$$\hat{\mu}(z) = \exp \left(\int_0^\infty (e^{izx} - 1) \nu(dx) + i\gamma_0 z \right),$$

ここで、 $\gamma_0 \in [0, \infty)$ で測度 ν はレヴィ測度と呼ばれ、 $[0, \infty)$ 上の $\nu(\{0\}) = 0$ と $\int_0^\infty (1 \wedge x) \nu(dx) < \infty$ を満たす。複合ポアソン分布では $\gamma_0 = 0$ で ν の全測度がポアソン分布の平均、 ν を全測度で割って正規化した確率分布が複合される確率分布になる。

無限分解可能分布とそのレヴィ測度の遠方での挙動の比較を考える。この問題には、分布の劣指数性 (Subexponentiality: $\lim_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\bar{F}(x)} = 2$ となる $[0, \infty)$ 上の分布で、べき法則など重い裾を意味する。* は分布の畳込み = 確率変数での独立和を表す。) と関連した有名な結果として、次の三つが同値であることが知られている (Embrechts et al. (1979))。 (i) 無限分解可能分布 μ が劣指数的である。 (ii) レヴィ測度 ν を正規化した分布が劣指数的である。 (iii) μ と ν の裾が漸近的に等しい ($\lim_{x \rightarrow \infty} \bar{\mu}(x) / \bar{\nu}(x) = 1$)。

このような無限分解可能分布とそのレヴィ測度の遠方での対応を、劣指数的分布以外に対しても考え、劣指数分布を拡張した

$$\limsup_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\bar{F}(x)} < \infty$$

で定義される \mathcal{O} (オー)-劣指数的分布 \mathcal{OS} を導入した。 \mathcal{O} -劣指数的分布には、劣指数的な分布の他、逆正規分布のような畳込み同値分布 (2 より大きい極限 $\lim_{x \rightarrow \infty} \overline{F * F}(x) / \bar{F}(x)$ をもつ) などがああり、この分布族に対して、次の定理が成り立つ。

定理 2 ([2])

- (i) $\mu \in \mathcal{OS} : \limsup_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\bar{F}(x)} < \infty,$
- (ii) ある自然数 n に対して $\nu_1^{n*} \in \mathcal{OS}$, ここで $\nu_1 = 1_{\{x > 1\}} \nu / \bar{\nu}(x)$.
- (iii) ある自然数 n に対して $\bar{\mu}(x) \asymp \bar{\nu}^{n*}(x)$:
($0 < \liminf_{x \rightarrow \infty} \bar{\mu}(x) / \bar{\nu}^{n*}(x) \leq \limsup_{x \rightarrow \infty} \bar{\mu}(x) / \bar{\nu}^{n*}(x) < \infty$).

現在、無限分解可能分布が劣指数性を持つ確率密度を持つ場合に対して、研究中である。劣指数的密度 (Subexponential density) の定義のみ示しておく。密度関数 $p(x)$ の裾が長い (long-tailed) とは、任意の実数 λ に対して、 $\lim_{x \rightarrow \infty} p(x + \lambda) / p(x) = 1$ を満たすときをいい、さらに、裾が長い密度関数 $p(x)$ が劣指数的とは $p^{2 \otimes}(x) \sim 2p(x)$ (\otimes は畳込み) を満たすときをいう。このような密度に対する仮定は考察対象を狭める半面、密度の漸近的性質は裾確率のそれよりも詳細な情報をもつ点で意義がある。

統計数理研究所では、以上の研究と関連した共同研究集会 「極値理論の工学への応用」と「無限分解可能過程に関連する諸問題」をそれぞれ夏と秋に開催しています。共同研究集会の情報は統数研ホームページイベント欄及び発表者のホームページ <https://sites.google.com/view/takaakishim> に掲載します。(事前登録・参加料などは必要ありません)。

参考文献

- [1] A numerical characteristic of extreme values, *Statistics and its Interface*.7(2014), 375-379.
- [2] Infinite divisibility and generalized subexponentiality, (with T. Watanabe (Univ. of Aizu)), *Bernoulli*, 11(2005), 445-469.