確率分布の上端点付近の挙動の研究

志村　隆彰　数理・推論研究系　准教授

【確率分布の遠方での性質】

ランダムな現象において、極端に大きい（或は小さい）事象は起こることが極めて稀にもかかわらず現実的にも理論（数学）的にも大きな意味を持つことがある。巨大地震や豪雨及びそれに伴う水害などの激甚災害はその典型例であり、最高気温に近いような酷暑やハイレベルなスポーツの記録までも同種のものとみなすことができる。このような場合、それらの極端事象そのものだけでなく、それがどの程度起こるかという頻度が重要である。ランダムな現象の起こり方は数学的に確率分布で表現されるが、極端現象を考察するには、確率分布全体ではなく、その一部である遠方での性質（正確には上端点近傍の挙動）のみを考えればよい。その際、よく用いられるのが、確率分の裾（確率）(tail probability) であり、分布 F に対して、

$$\overline{F}(x) = 1 - F(x)$$

で定義される。裾はxの関数として、x 以上の事象が起こる確率を表し、$x \to x_F := \sup\{x : F(x) < 1\}$ （x_F は F の上端点で、無限と有限の両方がありうる））のとき、単調（非増加）に 0 に近づく。同じxに対して、裾確率が大きいことはその確率分布が大きな値を取りやすいことを意味するから、$x \to x_F$　のときの 挙動(0へ行く程度（速さ））が重要となる。これが速いとき裾が軽い、遅いとき裾が重いという。裾が軽い分布と重い分布の典型例である、（標準）正規分布と（標準）コーシー分布を比べてみよう。それぞれの確率密度関数$p(t)$は、

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, \quad \frac{1}{\pi(1+x^2)}$$

であり、このふたつは原点に近いところでは一見よく似ているが、裾確率$\overline{F}(x) = \int_x^\infty p(t)dt$ の挙動は大きく違う。この違いは、たとえば、コーシー分布に従う独立同分布確率変数 $X_1, X_2, \ldots$ の算術平均の分布はひとつの確率変数の分布と等しくなって、正規分布と違って大数の法則が成り立たない。

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \stackrel{\mathrm{d}}{=} X_1 \quad \text{for each } n \in \mathbf{N}$$

裾が重いコーシー分布は極端に大きな値や小さな値をとることが多く、数学的意味での平均が存在せず、裾が軽い分布とは違った扱いが必要になることがわかる。裾挙動の重要性の一端をみたところで、具体的な問題を簡単に紹介する。

【極値理論】

極値理論は極端事象を扱う分野であり、伝統的には災害対策と結びついて、近年では金融を始めとするリスク管理全般の基礎として発展している。もっとも基本となるのは次の定理である。

$X_1, X_2, \ldots$ を共通の確率分布Fに従う実数値独立確率変数列、そのX_nまでの最大値を $M_n = \max\{X_1, \ldots, X_n\}$ とおき、この $n \to \infty$ のときの挙動を正規化して非退化分布に収束させる。適当な定数列$a_n > 0$と$b_n \in \mathbf{R}$により、

$$\frac{M_n - b_n}{a_n} \stackrel{\mathrm{d}}{\to} G \quad (n \to \infty).$$

このときの極限分布Gを極値分布といい、フレシェ分布、グンベル分布、（極値）ワイブル分布の３種類があり、それぞれの極値分布に収束する分布Fの全体はその（最大値）吸引領域と呼ばれ、裾（確率）$\overline{F}(x)$ の $x \to x_F$ のときの漸近挙動で特徴付けされる。理論的にはこれで完結しているが、このことを裾の軽い分布から生ずる実データに適用した場合、データの丸めなどの誤差が影響して、この極限定理が成り立ちづらいだけでなく、裾が重い場合であっても関連する推定量の性質などにも悪影響が生じることが知られている。

こうした場合への対処方法を見出すことが現在の研究目標である。指数分布のような指数的な裾を持つ分布に対しては、適当な補正をすることで誤差の影響を消す方法を見つけているが、それよりも軽い裾に対しては、対策が依然不明である。

【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。

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

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（$\lim_{x\to\infty} \overline{F * F}(x)/\overline{F}(x) = 2$ 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（$\lim_{x\to\infty} \overline{\mu}(x)/\overline{\nu}(x) = 1$）。

無限分解可能分布とそのレヴィ測度の遠方での性質に関する同種の問題は、劣指数的分布以外に対しても考えられている。

$[0,\infty)$ 上の分布FがO（オー）劣指数的($F \in \mathcal{OS}$)とは、

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

となるときをいい、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 $\lim_{x\to\infty} \overline{F * F}(x)/\overline{F}(x)$をもつ）などを含む分布族であり、これに対して、次の定理が成り立つ。

【定理(Shimura・Watanabe(2005))】

(i) $\mu \in \mathcal{OS}$,

(ii) ある自然数 n に対して $\nu_1^{n*} \in \mathcal{OS}$, ここで $\nu_1 = 1_{\{x>1\}}\nu/\overline{\nu}(x)$.

(iii) ある自然数 n に対して $\overline{\mu}(x) \asymp \overline{\nu^{n*}}(x)$:
$(0 < \liminf_{x\to\infty} \overline{\mu}(x)/\overline{\nu^{n*}}(x) \le \limsup_{x\to\infty} \overline{\mu}(x)/\overline{\nu^{n*}}(x) < \infty)$.

現在は、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、渡部俊朗教授（会津大学）とともに研究中である。密度の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。

局所劣指数性を始めとする多くの概念が必要なので、正確な結果は述べられないが、劣指数性を持つ確率密度を持つ無限分解可能分布とそのレヴィ測度の関係は、他の分布族に対する従来の結果と近い形ではあるものの、必ずしも単純に予想されるものとは一致しない微妙なものである。

最後に、以上の研究と関連した統計数理研究所共同利用事業である共同研究集会について紹介する。

【極値理論の工学への応用】

共同研究集会として「極値理論の工学への応用」を1994年以来毎年開催しています。今年度は来月７月１８日（木）、１９日（金）に開催ですので、興味のある方はご参加ください（事前登録・参加料などは必要ありません）。

【無限分解可能過程に関連する諸問題】

統数研共同研究集会「無限分解可能過程に関連する諸問題」を毎年開催しています。現時点では日程未定ですが、秋から初冬に開催します。

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http://www.ism.ac.jp/ shimura/ に掲載します。