万が一で起こることを考える－確率分布の裾に関する研究－

志村　隆彰　　数理・推論研究系 助教

1　万が一を考える

気象など自然に関するもの、株価など経済に関するものから、スポーツまで日々の生活の中で統計データを目にしない日はありません。そうした統計データは、簡単な数値によってある現象の性質を表すために使われますが、そうしたデータの性質を見てみると、大変よく用いられ、かつ対照的なふたつの種類のデータ、すなわち、ある現象の全体的傾向を表すものと極端な側面を表すものがあることに気がつくでしょう。例えば、天気予報における平均気温と最高（最低）気温、株価での平均株価と最高値（最安値)、平均寿命と最高齢記録、テレビの平均視聴率と瞬間最大視聴率、最近では、ネットショッピングである商品の平均価格と最安値といったものも容易に入手することが出来るようになっています。これら性質の異なる数値によって、我々はある現象を全体的な傾向とそのばらつき具合を知ることができるわけですが、同じ短距離走でも、子供の運動・運動能力調査では平均値が、五輪のような競技会では最高記録が問われるようにどちらのタイプの数値がより重要であるかはケースバイケースです。そして、確率分布の裾に関する研究というのは、ある現象の全体的な傾向ではなく、タイトルにもあるように、稀にしか起こらない極端な事象の起こり方についての研究ということになります。しかし、稀にしか起こらないから考える必然性が低いかといえば決してそうではありません。保険を考えれば、稀に起きるだけだけれども、一旦起きれば致命的であることは珍しくなく、安定した生活を送るために万が一に対する備えが必要なことは明白でしょう。

2　確率分布の裾

確率分布はランダムな現象においてどのような事象がどの程度起こるのかを表すものですが、確率分布の裾とは対応する確率変数がある値以上を取る確率のことです。確率変数をX、ある値をxとすれば、$Pr(X \geq x)$と書け、極端なことがらの起こりやすさを表します。裾確率 $Pr(X \geq x)$は xが大きくなるにつれ、0に近づきますが、だからといって意味がないわけではありません。近づき方が重要なのです。例えば、0 への近づき方が速いとき裾が軽い、遅いとき裾が重い（或いは厚い）といいます。軽いものとしては正規分布など、重いものとしてはパレート分布などがあります。大まかにいうと、裾が軽ければ、極端なことは無視して平均的なことを考えれば良く、逆に重ければ、平均を見るよりも極端なことを見なければならないといえます。歴史的には裾の軽い分布の研究が先行していて、確率変数の和に対する有名な理論である大数の法則や中心極限定理は裾が軽いときに成り立つ法則です。これに対し、重い分布の世界では中心極限定理が成り立ちません。多くの確率変数の独立和がそのうちの最大のものたったひとつとほとんど同じであるということも起こりえます。近年、重い裾に基づく現象の発見及びそれに伴う応用上の必要性、加えて解析手法としての数学の進歩により裾の重い分布の研究が盛んになってきました。

3　極値統計学

裾の挙動が主役となる分野の代表が極値統計学です。極値統計学、極値理論では、データの和ではなく、データの中で最大のものの挙動などが研究の対象になります。例えば、洪水対策で堤防の設計をする際には、わずかな雨量の日数の多さよりも、たとえ日数は僅かであっても大雨の頻度が重要です。このように、極値理論は数多く起こる小さなものの蓄積よりも稀に起こる極端に大きいことが意味を持つ現象の解析に用いられます。極値理論は、自然災害の分野だけでなく、建築工学、信頼工学、保険数学、ファイナンスなどの安全性やリスクを扱う工学の分野で広く用いられています。

中心極限定理に対応する極値理論の基本定理として次が知られています：$X_1, X_2, \ldots$ 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$M_n = \max\{X_1, \ldots, X_n\}$ とおき、適当な定数列 A_n と B_nにより、

$$\mathcal{L}\left(\frac{M_n - A_n}{B_n}\right) \to G \quad (n \to \infty).$$

となるとき、極限分布Gを極値分布といい、フレシェ分布、（逆）ワイブル分布、グンベル分布の3種類があります。標本分布が指数分布や正規分布であれば、グンベル分布に収束します。

4　無限分解可能分布

平均 m、分散 v の正規分布はどのような自然数 n に対しても平均 m/n、分散 v/n の正規分布の n回合成積（確率変数で言えばn個の独立同分布の列の和）となります。このようにどの自然数 n に対してもn個の同じ分布のn回合成積に分解できる分布を無限分解可能分布といいます。ポアソン分布、複合ポアソン分布は無限分解可能ですが、一様分布、二項分布はそうではありません。無限分解可能分布は基本的な確率分布として対応する確率過程ともども古くから理論的研究がなされてきました。裾が重い分布としてよく知られている安定分布は代表的な無限分解可能分布で、これをはじめとして、最近ではファイナンスのモデルなどに応用されることが多くなってきています。

無限分解可能分布の特性関数（フーリエ変換）は一般に次のような形をしています。

$$\int_{-\infty}^{+\infty} e^{izt}F(dt) = \exp\{iaz - \frac{1}{2}vz^2 + \int_{-\infty}^{+\infty}(e^{izt} - 1 - \frac{izt}{1+t^2})G(dt)\},$$

ここで、$a \in \mathbf{R}$、$v \geq 0$、G は一定の条件を満たす測度でレヴィ測度と呼ばれます。

無限分解可能分布分解 F とそのレヴィ測度Gの裾同士の関係は出力と入力との関係と見なすことが出来、これを探ることは重要な問題です。

5　正則変動関数

裾確率の挙動を数学的に記述するのに正則変動関数がよく用いられます。

正値可測関数 $f(x)$ が（∞ で）指数 $\rho(\in \mathbf{R})$ の正則変動関数であるとは、任意の $k > 0$ に対して、

$$\lim_{x\to\infty} f(kx)/f(x) = k^\rho$$

を満たすときをいいます。特に指数 0 のときを緩慢変動関数といい、指数 ρ の正則変動関数は x^ρ と緩慢変動関数の積になります。緩慢変動関数は徐々に変動が小さくなっていく関数で、次のようにいろいろなものがあります。$l_1(x) = c + o(1)$. ここで、c は正の定数、$l_2(x) = \log x$、$l_3(x) = (\log x)^2$、$l_4(x) = \exp\{(\log x)^{\frac{1}{3}}\cos((\log x)^{\frac{1}{3}})\}$。

例えば、裾が正則変動する分布に従う独立同分布列では、独立和を正規化したものが安定分布へ、最大値を正規化したものがフレシェ分布へ、それぞれ収束することが知られています。

定義からわかるように正則変動関数は多項式（べき乗関数)オーダーを滑らかに拡張した概念で、これにより一層精密な扱いが可能になります。現在では様々な派生概念も生まれており、統計・数学における多くの分野で活用されています。

6　関連活動

統計数理研究所では、以下のような一般対象の講座（要事前申込、有料）や研究者対象の研究集会を行っています。詳細は統計数理研究所ホームページ（http://www.ism.ac.jp/index.html）或いは
志村（http://www.ism.ac.jp/ shimura/）をご覧ください。

公開講座 「極値統計学」（8月25日（水）、26日（木））7月20日（火）10時から募集開始、先着順40名。

共同研究集会 「極値理論の工学への応用」（秋開催予定）

共同研究集会 「無限分解可能過程に関連する諸問題」（秋開催予定）