極値データの特徴

志村　隆彰　数理・推論研究系　助教

【極値統計学とは】

地震や洪水などの自然災害に代表される、めったに起こらないが、一旦起こると大変大きな影響を及ぼしたり、重要な意味を持つ現象は数多い。このような現象を統計的に扱う場合、日常を表す平均値のような指標ではなく、非日常を表わす最大値のような指標が重要になる。極値統計学はいわば非日常を研究対象にする分野であり、数学的にもっとも簡単で基本的な設定と関心事は $X_1, X_2, \ldots$ を共通の確率分布 F に従う実数値独立確率変数列としたときの X_n までの最大値 $M_n = \max\{X_1, \ldots, X_n\}$ の $n \to \infty$ のときの挙動である。M_n が F の上端点 $x_F = \sup\{x : F(x) < 1\}$（無限と有限の両方がありうる）に収束するのは明らかであるから、その収束の速さや正規化したときの極限分布（極値分布）を求めるのが基本的な問題であり、裾（確率）

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

の $x \to x_F$ のときの挙動が重要である。

F の上端点が無限であれば、M_n も無限に行き、それは大きなランダムな数である。この発表では大きなランダムな数値の数値的特徴について述べる。

【大きなランダムな数の正規化】

大きなランダムな数の先頭から二つ目以降の数の様子を見たい。そのために適当な正規化を施す。X を上限無限の分布 F に従う確率変数とし、X を一桁($[1, 10)$)の数にスケール変換する。正確には、N を X の桁数、K を X の先頭の数とし、$Y = (X - K10^{N-1})/10^{N-1}$ とするとき、次の条件付き確率分布

$$F^{k,n}(x) = P(Y \le x | K = k, N = n),$$
$$k = 1, 2, \ldots, 9.$$

を考える。主な関心は $n \to \infty$ のときの $F^{k,n}$ の極限分布である。

結果を述べるために必要な正則変動関数及び関連事項を説明する。

【正則変動性に関連する関数】

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る。これは多項式オーダー（べき乗関数）を滑らかに拡張した概念で以下のように定義される。

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

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

を満たすときをいう。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 $l(x)$ ($\lim_{x\to\infty} l(\lambda x)/l(x) = 1$) の積になる。指数が 0 でなければ正則変動関数は漸近的な意味で単調（指数正なら増加、負なら減少）になる。一方、指数 0 の緩慢変動の場合は、自明な $c + o(1)(c > 0)$、対数関数 $\log x$、振動するもの $\exp\{(\log x)^{\frac{1}{3}} \cos((\log x)^{\frac{1}{3}})\}$ まで挙動は多様だが、単調なものが現れることがあり、中でも重要なのが、Π 変動関数である。

非減少正値可測関数 $f(x)$ が のΠ変動関数であるとは、$a(x) > 0, b(x)$ が取れて、各 $\lambda > 0$ に対して、

$$\lim_{x\to\infty} (f(\lambda x) - b(x))/a(x) = \log \lambda$$

を満たすときをいう。対数関数などが Π変動するが（$a(x) = 1$、$b(x) = \log x$)、この性質は局所的な滑らかさを要するため、漸近的に閉じてはいない。

緩慢変動関数とは逆に指数を無限にしたものを（指数無限の）急速変動関数と呼び、任意の $\lambda > 1$ に対して、$\lim_{x\to\infty} f(\lambda x)/f(x) = \infty$ で定義される。指数関数などはこれに当たる。指数無限の急速変動関数の逆数を指数マイナス無限の急速変動関数と言う。それぞれ $\mathbf{R}_\infty$、$\mathbf{R}_{-\infty}$ で表わす。

【極値統計量の数値的特徴】

上述の関数により裾の挙動の特徴付けられる分布の例を見よう。まず裾が急速変動する分布の例は、裾が軽い方から正規分布、レイリー分布などの非常に早く裾が減衰するもの、指数的裾を持つといわれる、指数分布、ガンマ分布、カイ二乗分布、一般化逆ガウス分布など、これらより裾が重い対数正規分布など数多くある。次に裾が（指数が負の）正則変動するものには、コーシー分布を始めとする安定分布、パレート分布、F分布、ジップ分布などがある。更に裾が重いものとして、対数コーシー分布(対数をとればコーシー分布）を考えれば、裾の逆数が Π 変動する。

そして、$F^{k,n}$ の $n \to \infty$ のときの極限分布 F^k は F がどの分布族に属するかによって以下のように決まる。

定理１

(i) $\bar{F}(x) \in \mathbf{R}_{-\infty}$ ならば、全ての k に対して、

$$\lim_{n\to\infty} F^{k,n}(x) = 1_{\{x \ge 1\}},$$

ここで、1_A は集合 A の指示関数。

(ii) $\bar{F}(x) \in \mathbf{R}_{-\alpha}(\alpha > \mathbf{0})$ ならば、$0 \le x \le 1$ に対し、

$$F^k(x) = \lim_{n\to\infty} F^{k,n}(x) = \frac{1 - (1 + \frac{x}{k})^{-\alpha}}{1 - (1 + \frac{1}{k})^{-\alpha}}.$$

(iii) $1/\bar{F}(x) \in \Pi$ ならば、$0 \le x \le 1$ に対し、

$$F^k(x) = \lim_{n\to\infty} F^{k,n}(x) = \frac{\log(1 + \frac{x}{k})}{\log(1 + \frac{1}{k})}.$$

(iii) は他と違う形になっているが、実際次が成り立つ。

定理２　裾が緩慢変動する任意の分布 F と任意の $[0, 1)$ の分布 G に対して、$\lim_{x\to\infty} \bar{F}_G(x)/\bar{F}(x) = 1$ かつ $F_G^{k,n} = G$ (n と k によらない) となる F_G が存在する。

定理１ (i) における δ_0 分布への収束の速さに関して以下が成り立つ。

定理３　$\bar{F}(x) \in \mathbf{R}_{-\infty}$ は絶対連続でそのハザード関数 $h(t)$ が $\mathbf{R}_\rho(\rho \ge -\mathbf{1})$ に入ると仮定する。このとき、$0 \le x < 1$ に対し、

$$\lim_{n\to\infty} \frac{1}{10^{n-1} h(10^{n-1})} \log \overline{F^{k,n}}(x) = -c(\rho, k, x),$$

ここで、

$$c(\rho, k, x) = \begin{cases} (\rho + 1)^{-1}\{(k + x)^{\rho+1} - k^{\rho+1}\} & \rho > -1 \text{のとき} \\ \log(1 + \frac{x}{k}) & \rho = -1 \text{のとき} \end{cases}$$

【極限分布の性質】

定理１から裾が正則変動する分布に従う大きな数値では、先頭の数字と二つ目以降の数に従属性があるからこれを極限分布の平均値を通じて見てみよう。

定理４　定理１の (ii) と (iii) の極限分布 F^k の平均 M_k は以下の通り。

$$M_k = \begin{cases} \frac{(\alpha+k)(k+1)^{-\alpha} - k^{-\alpha+1}}{(1-\alpha)(k^{-\alpha} - (k+1)^{-\alpha})} & \alpha \ne 1 \text{のとき} \\ k(k+1)\log(1 + \frac{1}{k}) - k & \alpha = 1 \text{のとき} \\ 1 - k\log(1 + \frac{1}{k}) & 1/\bar{F}(x) \in \Pi \text{のとき} \end{cases}$$

いずれの場合も、平均値 M_k は k に対して増加する。

http://www.ism.ac.jp/ shimura/