2014年12月8日

極値分布の確率論的な基礎知識
ー裾の挙動から見た確率分布ー

志村 隆彰

（統計数理研究所）

【確率分布の裾】

極値統計が対象としている極端な事象の起こり方は分布の裾と呼ばれる次で与えらえる.

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

ここで $F(x)$ は確率分布 F の分布関数である.

【確率分布の上限点】

$$x_+(F) \equiv \sup\{x : F(x) < 1\} (\leq \infty)$$

を 分布 F の上限点という.

x が F の上限点 $x_+(F)$ へ近づけば，裾 $\bar{F}(x)$ は 0 に近づく.

$$\lim_{x \to x_+(F)} \bar{F}(x) = 0.$$

このときの0への近づき方（速さ）は分布により様々であり、ランダムな現象を決定づける.

【様々な確率分布の裾の挙動】

代表的な確率分布の裾 $\bar{F}(x)$ の $x \to \infty$ のときの挙動を挙げる.

(i) 標準正規分布 $\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}$

(ii) 標準コーシー分布 $\frac{1}{\pi} \frac{1}{x}$

(iii) 指数分布 $e^{-\lambda x} \quad (\lambda > 0)$

(iv) パレート分布 $x^{-\alpha} (\alpha > 0)$

このような裾の減衰の速さを表すために正則変動関数とその派生概念が用いられる.

【講演目的】
この講演では正則変動関数がどのような関数で，極値統計（理論）でどのように用いられるのかを紹介する．

キーワード：　確率分布の裾，正則変動関数，最大値吸引領域

講演項目：

1. 正則変動関数とその拡張

1．1 正則変動関数の定義と例

1.2. 基本的な性質と重要な定理

1.3. 正則変動関数の派生概念

1.4. 大域的性質と一般化逆関数

2. 正則変動関数の応用-吸引領域の特徴付け

2.1. 極値分布の吸引領域

2.2. 吸引領域の特徴付けと例

1　正則変動関数概論

正則変動関数とは，$x \to \infty$ のときの漸近挙動がべき乗関数 x^ρ に “近い” 関数である．正確には，べき乗関数に緩慢変動関数と呼ばれる，ゆっくりと変化する関数 $l(x)$ をかけたものである．

$$x^\rho \times l(x)$$

極値統計（理論）において確率分布の裾の挙動の特徴付けなどに用いられる．

1.1　正則変動関数の定義と例

【正則変動関数】

$[x_0, \infty)$ 上の正値可測関数 $f(x)$ が（∞ で）指数 ρ の正則変動関数である $(f \in \mathbf{RV}_\rho)$ とは，任意の $\lambda > 0$ に対して，

$$\lim_{x\to\infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho \tag{1}$$

を満たすときをいう．特に，指数 0，すなわち (1) の右辺が 1 のとき，緩慢変動関数と呼ぶ．

指数 ρ の正則変動関数は，べき乗関数 x^{ρ} と緩慢変動関数 $l(x)$ の積で表される.

$$f(x) = x^{\rho} l(x), \quad \lim_{x \to \infty} \frac{l(\lambda x)}{l(x)} = 1 \qquad (2)$$

たとえば、$l(x) = \log x$ ならば、

$$\frac{\log \lambda x}{\log x} = \frac{\log x + \log \lambda}{\log x} \to 1 \ (x \to \infty)$$

記号：$\mathbf{RV} = \cup_{\rho \in \mathbf{R}} \mathbf{RV}_{\rho}$.

【緩慢変動関数(1-5) と正則変動しない例(6,7)】

1. $l(x) = c + o(1)$. ここで，c は正の定数，$\lim_{x\to\infty} o(1) = 0$.
2. $l(x) = \log x$.
3. $l(x) = (\log x)^2$.
4. $l(x) = \exp\sqrt{\log x}$.
5. $l(x) = \exp\{(\log x)^{\frac{1}{3}} \cos((\log x)^{\frac{1}{3}})\}$.
6. $l(x) = e^x$.
7. $l(x) = 2 + \sin x$.

1.2　基本的な性質と重要な定理

【命題1.1】$l(x)$，$l_1(x)$，$l_2(x) \in \mathbf{RV_0}$ とする.

(i) $\alpha \in R$ に対して，$l(x)^\alpha \in \mathbf{RV_0}$.

(ii) 積 $l_1(x)l_2(x)$，和 $l_1(x)+l_2(x)$，さらに $\lim_{x\to\infty} l_2(x) = \infty$ ならば，合成関数 $l_1(l_2(x))$ も緩慢変動関数である.

(iii) 任意の $\varepsilon > 0$ に対し，$\lim_{x\to\infty} x^{\varepsilon} l(x) = \infty$ $\lim_{x\to\infty} x^{-\varepsilon} l(x) = 0$.

(iv) $\lim_{x\to\infty} \log l(x)/\log x = 0$.

【定理1.1（一様収束定理)】
緩慢変動関数 $l(x)$ の収束，

$$\lim_{x \to \infty} \frac{l(\lambda x)}{l(x)} = 1$$

は，$\lambda \in (0, \infty)$ について広義一様収束するすなわち、$l(x)$ が緩慢変動関数ならば，任意の小さい $\epsilon > 0$ に対して，

$$\lim_{x \to \infty} \sup_{\epsilon \leq \lambda \leq 1} \left| \frac{l(\lambda x)}{l(x)} - 1 \right| = 0. \tag{3}$$

1.3　派生概念-Γ変動性とΠ変動性

【定義1.3】$(x_0, x_+)(x_0 < x_+ \leq \infty)$ 上の非減少関数 $f(x)$ が Γ 変動 $(f \in \Gamma)$ するとは，$\lim_{x \uparrow x_+} f(x) = \infty$，$(x_0, x_+)$ 上の正値関数 $r(x)$ があって，任意の λ に対して，

$$\lim_{x \to x_+} \frac{f(x + \lambda r(x))}{f(x)} = e^{\lambda}$$

となるときをいう．$r(x)$ を補助関数(auxiliary function)といい，漸近的な意味で一意に決まる．

$$\lim_{x\to\infty} \frac{\exp(\rho(x+\lambda\rho^{-1}))}{\exp(\rho x)} = e^{\lambda}$$

【例】

1. $f(x) = e^{\rho x}$ $(x_+ = \infty)$ は補助関数 $r(x) = \rho^{-1}$ のΓ変動関数である.
2. $f(x) = \exp\left(1/(1-x)\right)(0 < x < 1)$ は補助関数 $r(x) = (1-x)^2$ のΓ変動関数である.
3. $f(x) = \exp[x]$ は，Γ変動関数ではない.

【定義1.4】半直線(x_0, ∞)上の非負非減少関数$g(x)$がΠ変動（$U \in \Pi$）するとは，関数$a(x) > 0$と $b(x) \in \mathbf{R}$ があって，任意の $\lambda > 0$ に対して，

$$\lim_{x \to \infty} \frac{g(\lambda x) - b(x)}{a(x)} = \log \lambda.$$

となるときをいう．$a(x)$ を補助関数といい，漸近的な意味で一意に決まる．

$b(x) = g(x)$, $a(x) = g(ex) - g(x)$ としてよい.

$$\lim_{x\to\infty} \frac{g(\lambda x) - g(x)}{g(ex) - g(x)} = \log \lambda.$$

【例】

1. $g(x) = \log x$ は $a(x) = 1$ のΠ変動関数.
2. $g(x) = 1 - (\log x)^{-1} (x \geq e)$ は補助関数 $a(x) = (\log x)^{-2}$ のΠ変動関数である.
3. $g(x) = [\log x]$ は非減少緩慢であるが，Π変動関数ではない.

1.4　正則変動関数の大域的性質と一般化逆関数

緩慢変動の場合とは異なり，指数が0でない正則変動関数は指数の符号に応じて，無限大と０へ行くが，そのときの挙動は漸近的に単調になる．この性質により、一般化された逆関数を考えることが出来る.

更に、緩慢変動関数の場合も、Π変動するものに対しては同様である.

【定理1.4（Monotone equivalence）】

$f(x) \in \mathbf{RV}_\rho$ を $[x_0, \infty)$ $(x_0 \geq 0)$ 上で局所有界とする．$\rho > 0$ ならば,

$$
\begin{aligned}
(i) \quad & \overline{f}(x) \equiv \sup\{f(t) : x_0 \leq t \leq x\} \sim f(x), \\
(ii) \quad & \underline{f}(x) \equiv \inf\{f(t) : t \geq x\} \sim f(x).
\end{aligned}
$$

$\rho < 0$ ならば,

$$
\begin{aligned}
(i) \quad & \sup\{f(t) : t \geq x\} \sim f(x), \\
(ii) \quad & \inf\{f(t) : x_0 \leq t \leq x\} \sim f(x).
\end{aligned}
$$

【定理1.5（漸近的逆関数）】$f(x) \in \mathbf{RV}_\alpha$ $(\alpha > 0)$ ならば,

$$f(g(x)) \sim g(f(x)) \sim x$$

を満たす $g(x) \in \mathbf{RV}_{\mathbf{1}/\alpha}$ が存在する．この $g(x)$ を $f(x)$ の漸近的逆関数$(asymptotic\ inverse)$ と呼び，漸近的に唯ひとつに決まる．具体的に，以下の一般化された逆関数をとることが出来る．

$$f^{\leftarrow}(x) \equiv \inf\{y \in [x_0, \infty) : f(y) > x\}.$$

【例】

1. $f(x) = x^\rho$ と $g(x) = x^{1/\rho}$ は互いに（漸近的）逆関数である.
2. $f(x) = (\log x)x$ と $g(x) = (\log x)^{-1}x$ は互いに漸近的逆関数である.

次は指数が０でない正則変動関数間の逆関数の関係を拡張したものである.

【定理1.6】

(i) $f(x) \in \Gamma$でその補助関数が$r(x)$ならば,
$f^{\leftarrow}(x) \in \Pi$であり，補助関数は
$a(x) = r \circ f^{\leftarrow}(x)$である.

(ii) $g(x) \in \Pi$でその補助関数が$a(x)$ならば,
$g^{\leftarrow}(x) \in \Gamma$であり，補助関数は
$r(x) = a \circ g^{\leftarrow}(x)$である.

【例】

1. $f(x) = e^x$ と $g(x) = \log x$ は互いに逆関数である.
2. $f(x) = \exp\frac{1}{1-x}$ $(0 \leq x < 1)$ と $g(x) = 1 - (\log x)^{-1}$ は互いに逆関数である.

2　正則変動関数の応用—最大値吸引領域の特徴付け

極値理論において正則変動関数が用いられる代表的な例として，最大値吸引領域を挙げて解説する．

2.1　極値分布の吸引領域

【定義2.1】$X_1, X_2, \ldots$ 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

$$M_n = \max\{X_1, \ldots, X_n\}$$

とおく．M_n を定数列 a_n，b_n により正規化し，$n \to \infty$ としたとき，非退化分布 G へ収束するならば，極限の分布 G を極値分布という．

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

このとき，F は G に吸引されるといい，そのような分布 F の全体を G の（最大値）吸引領域（$D(G)$ と書く）とよぶ(Maximum domain of attraction).

【定理 [Fisher-Tippett の定理]】

極値分布は次の３種のいずれかである．それぞれ，フレシェ（Fréchet）分布，（負の）ワイブル（Weibull）分布，グンベル（Gumbel）分布と呼ばれる．以下，$\alpha > 0$ とする．

(i)

$$\Phi_\alpha(x) = \begin{cases} 0 & (x \leq 0) \\ \exp(-x^{-\alpha}) & (x > 0) \end{cases}$$

(ii)

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & (x \le 0) \\ 1 & (x > 0) \end{cases}$$

(iii) $$\Lambda(x) = \exp(-e^{-x}).$$

2.2　各極値分布の吸引領域の特徴付け

吸引領域への属性は分布の裾

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

の F の上限点

$$x_+(F) \equiv \sup\{x : F(x) < 1\}$$

付近の挙動で決まる.

そして、それは正則変動性で表現される.

“多くの種類の分布，とりわけほとんどの連続分布は，いずれかの極値分布の吸引領域に入る”.

各極値分布の吸引領域の特徴として

$D(\Phi_\alpha)$ に属する分布は上限点が無限，すなわちいくらでも大きい値をとりうる.
$D(\Psi_\alpha)$ に属する分布の上限点は有限，すなわち取る値に限界がある.
$D(\Lambda)$ に属する分布はその両方の場合がある.

では，３種の極値分布それぞれの吸引領域に入るための必要十分条件について述べる．
それぞれの吸引領域は上限点での挙動が

$D(\Phi_\alpha)$ は正則変動性で特徴づけられる．
$D(\Psi_\alpha)$ は上限点近傍での正則変動性で特徴づけられる．
$D(\Lambda)$ に属する分布は Γ 変動性で特徴づけられる．

【定理】[フレシェ分布の吸引領域]
$F \in D(\Phi_\alpha)$ $(\alpha > 0)$ となるための必要十分条件は $\bar{F}(x) \in \mathbf{RV}_{-\alpha}$ である.

正規化定数は,

$$a_n = \inf\{x : \bar{F}(x) \leq 1/n\}, \ b_n = 0 \qquad (4)$$

で定めることができる.

【$D(\Phi_\alpha)$ に属する分布の例】

コーシー分布は $D(\Phi_1)$ に含まれる.

$$p(t) = \frac{1}{\pi(1+t^2)}$$

この他，パレート分布，ブール分布，非正規安定分布，対数ガンマ分布，ジップ分布（離散パレート分布）

【定理】[ワイブル分布の吸引領域]
$F \in D(\Psi_\alpha)$ $(\alpha > 0)$ となるための必要十分条件は$x_+(F) = \sup\{x : F(x) < 1\}$（$F$の上限点）が有限かつ$\bar{F}(x_+ - 1/x) \in \mathbf{RV}_{-\alpha}$である.
正規化定数は,
$a_n = \sup\{x : \bar{F}(x_+ - x) \leq 1/n\}$, $b_n = x_+$ と定めることができる.

【$D(\Psi_\alpha)$ に属する分布の例】
$(0, 1)$上の一様分布，他に，ベータ分布，（負の）ワイブル分布など

【定理】[グンベル分布の吸引領域]

$F \in D(\Lambda)$ となるための必要十分条件は,

(i) $1/\bar{F}(x) \in \Gamma$，すなわち，ある正値関数 $r(x)$ に対して,

$$\lim_{x \uparrow x_+} \frac{\bar{F}(x+\lambda r(x))}{\bar{F}(x)} = e^{-\lambda}.$$

$r(x) = \int_x^{x_+} \bar{F}(t)dt/\bar{F}(x),\ (t < x_+)$ ととることができる.

(ii) $U(x) = (1/\bar{F}(x))^{\leftarrow} \in \Pi$，すなわち，任意の $\lambda, \nu > 0$, $\nu \neq 1$ に対して，

$$\lim_{x\to\infty} \frac{U(\lambda x) - U(x)}{U(\nu x) - U(x)} = \frac{\log \lambda}{\log \nu}.$$

$$H(x) = -\log \bar{F}(x),$$

とし，その(右連続)逆関数を

$H^{\leftarrow}(x) = \inf\{t : H(t) \geq x\}$ とすれば，

正則化定数は

$$\begin{aligned} b_n &= H^{\leftarrow}(\log n), \\ a_n &= H^{\leftarrow}(\log(ne)) - H^{\leftarrow}(\log n) \end{aligned}$$

ととれる．

【$D(\Lambda)$ に属する分布の例】

(i) 上限点無限　指数分布.

(ii) 上限点有限

$$\bar{F}(x) = \exp\left(\frac{-1}{1-x}\right) \quad (0 \leq x < 1).$$

他に、正規分布，ガンマ分布，対数正規分布，（通常の）ワイブル分布など

【極値分布の吸引領域に入らない分布】

二項分布，幾何分布，ポアソン分布など.

上限のある離散分布，裾が重くない離散分布や極端に重い裾を持つ分布（スーパーヘヴィテール）などがどの極値分布の吸引領域にも入らない.

吸引領域の間には次のような対応がある.

【定理】X を上限無限の正値確率変数とする（$-1/X$ は上限０の負値確率変数）．このとき，X の分布が吸引領域に属することと $-1/X$ の分布が吸引領域に属することは同値となる.

【系】負の対数正規分布（対数正規分布を反転させた上限有限の分布）はグンベル分布の吸引領域に属する.

【まとめ】

· 極値理論では確率分布の裾の性質（減衰の速さ）が重要である.

· 確率分布を裾の視点から見ると普段とは違った分布の姿が見えてくる.

· 分布の裾の減衰の速さは様々で数学的には正則変動性で記述される.

· 各極値分布の吸引領域の特徴付けと具体例を挙げた.

ご清聴ありがとうございました