統計数理（2004）
第 52 巻 第 1 号 25–43

特集「極値理論」
［総合報告］

弱従属性をもつデータに基づく極値統計の最近の話題

吉原 健一[†]

（受付 2003 年 8 月 1 日；改訂 2003 年 10 月 9 日）

要　　旨

極値理論は理論的にみても実用的にみても，非常に重要な問題を数多く含んでいて，以前から盛んに研究され，現在も続いている．一方，現実の問題を解決するためには，通常使われている「標本は独立である」という仮定は条件として適当でない場合も多い．そこで，本総合報告では，標本がある種の弱従属性をもつ場合,「独立性を仮定して得られている結果はそのまま成り立つのか」または「少し修正すれば似たような結果は得られるのか」，さらにまた,「独自の結果が出てくるのか」を，最近のいくつかの論文を通して調べる．

キーワード： 弱従属性，指標，Hill 推定量，定常確率変数列.

1. はじめに

$\{\xi_n; n \geq 1\}$ を確率空間$(\Omega, \mathbf{F}, P)$の上で定義された強定常確率変数列で，共通の分布関数 $F(x) = P(\xi_n \leq x)$ をもつものとする．以下強定常確率変数列を単に定常確率変数列という．$\xi_a, \cdots, \xi_b (a \leq b)$ で生成される σ-集合体を $\mathbf{M}_a^b = \sigma(\xi_a, \cdots, \xi_b)$ で表し，

$$\alpha(n) = \sup_{k \geq 1} \sup_{A \in \mathbf{M}_1^k, B \in \mathbf{M}_{n+k}^\infty} |P(AB) - P(A)P(B)|$$

とおく．もし

$$\alpha(n) \to 0 \quad (n \to \infty)$$

が成り立てば，確率変数列 $\{\xi_n; n \geq 1\}$ は強混合型条件を満足するといい，この $\alpha(n)$ を強混合型条件の係数という．また，特に

$$\alpha(n) = 0 \quad (n \geq m+1)$$

が成り立つとき，この確率変数列は m-従属であるという．

これは，Rosenblatt (1956)が提唱した弱従属の概念である．この概念はもちろん独立の概念の拡張であり，$\alpha(n)$ の収束の早さに関係するが，この条件をみたせば，$\{\xi_n; n \geq 1\}$ が独立，同分布の場合に得られる結果は，ほとんどすべての場合に成立することが証明されている．当然，例外はある．

そこで，本稿では，独立な場合に得られている極値統計に関する結果は，どのように弱従属の場合に拡張されているのか，さらに，拡張されたものはどのように応用できるかを吟味した

[†] 創価大学 教育学部：〒192–0003 東京都八王子市丹木町 1–236

いと思う．

以下，第 2 節では Poisson like 確率変数の最大値の極限分布，第 3 節では，強混合型確率過程の極値の性質について，応用例（治癒過程）も含めて考察した．第 4 節では，Hill 推定量を中心とした極値指標の推定の問題を扱った．なお，これらの中で，独立性を仮定した結果も多くは弱従属性をもつ場合に拡張できる可能性の強いものである．

ここで，以下で使う記号の主なものを挙げる．任意の確率変数 Z の分布関数を $\mathcal{L}(Z)$ とかく．$\xi_1, \xi_2, \cdots$ が独立で同分布に従うときを i.i.d. で表すこととする．また，特に断らないときは ξ_i は実数値確率変数とする．$\mathcal{C}[0,1]$ は区間 $[0,1]$ の上で定義された連続関数の全体，$\mathcal{D}[0,1]$ は右連続かつ左極限をもつ区間 $[0,1]$ の上で定義された関数の全体とし，$[0,1]$ の上の値をとる確率変数列またはその分布列が $\mathcal{D}[0,1]$ に値をとる確率変数またはその分布に弱収束することを $\to^D$ in $\mathcal{D}[0,1]$，概収束することを $\to^{\mathrm{a.s}}$ とかく．

$[x]$ は x の整数部分，$\mathsf{R} = (-\infty, \infty)$, $\mathsf{R}_+ = [0, \infty)$，さらに，任意の空間 S に対しその Borel 集合の全体を $\mathrm{B}(\mathsf{S})$ で表す．

確率変数 X の分布関数 $F(x)$ の台が $[0,\infty)$ で

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\gamma}, \quad x > 0$$

が成り立つとき，尾部 $1-F(x)$ は指標 $-\gamma$ で正則に変動するといい，$1-F(x) \in RV_{-\gamma}$ とかく．一般に，H の逆関数は

$$H^{-1}(y) = \inf\{x | H(x) \geq y\} \quad (0 < y < 1)$$

によって定義する．そのとき

$$b(t) = \left(\frac{1}{1-F}\right)^{-1}(t) = F^{-1}\left(1 - \frac{1}{t}\right)$$

とおくと，$1-F \in RV_{-\gamma}$ ならば

$$1 - F(b(t)) \sim t^{-1}$$

が成り立つ．

さらに，任意の確率変数列 $\{\xi_n\}$ に対して，$\xi_1, \cdots, \xi_n$ の最大値を M_n とかく，すなわち

$$M_n = \max_{1\leq i\leq n} \xi_i \quad (n \geq 1).$$

$\{M_n\}$ の正規化数列として数列 $\{a_n\}$，$\{b_n\}$ を使うときは，$\{a_n\}$ はつねに正数列，$\{b_n\}$ は一般の数列とする．また，$n \to \infty$ のとき $h_n \to \infty$ かつ $h_n/n \to 0$ を満たす自然数列全体を K で表す．

2. 極値の極限分布

2.1 $\{\xi_n\}$ が i.i.d. の場合の M_n の極限分布

ξ_1 の分布関数を F とする．正規化数列 $\{a_n\}$，$\{b_n\}$ を適当に選べば退化しないある分布関数 G に対して

$$\mathcal{L}\left(\frac{M_n - b_n}{a_n}\right) \to G \tag{2.1}$$

となるとき，ある γ $(\in (0,\infty))$ に対して

$$\lim_{n\to\infty} n(1 - F(u_n)) = \gamma \tag{2.2}$$

を満たす数列 $\{u_n\}$ が存在することはよく知られている．$\{\xi_n\}$ が条件(2.1)を満たすとき，F は G の吸引域に属するといい，$F \in \mathsf{D}(G)$ とかく．(2.1)で得られる極限分布は次の 3 種のタイプの 1 つであることは知られている：

(I) $G_0(x) = \exp(-e^{-x}) \quad (-\infty < x < \infty);$

(II) $G_{-\gamma}(x) = \exp(-x^{-\gamma})\ (x \geq 0), \quad = 0\ (x < 0);$

(III) $G_{\gamma}(x) = 1\ (x \geq 0), \quad = \exp(-(-x)^{\gamma})\ (x < 0)$

ただし，$\gamma > 0$ である．

さらに，

$$x_F = \sup\{x|F(x) < 1\},$$
$$d_n = \inf\{x|F(x) \geq 1 - n^{-1}\},$$
$$g(t) = \int_t^{x_F} \frac{1 - F(u)}{1 - F(t)} du \quad (t < x_F)$$

とおくと，正規化数列 $\{a_n\}$，$\{b_n\}$ を次のようにとることができる：

$F \in \mathsf{D}(G_0)$ の場合 $\quad a_n = g(d_n),\ b_n = d_n$

$F \in \mathsf{D}(G_{-\gamma})$ の場合 $\quad a_n = d_n,\ b_n = 0$

$F \in \mathsf{D}(G_{\gamma})$ の場合 $\quad b_n = (x_F - d_n)^{-1},\ b_n = x_F.$

ここで数列 $\{v_n\}$ を

$$v_1 = E\xi_1;\ v_{n+1} = E(\xi_1 \vee v_n)\ (n \geq 1)$$

によって定義する．Kennedy and Kertz (1991)は次の定理を証明した．

定理 2.1. $F \in \mathsf{D}(G_0)$，$F \in \mathsf{D}(G_{-\gamma})\ (\gamma > 1)$ または $F \in \mathsf{D}(G_{\gamma})\ (\gamma > 0)$ ならば，それぞれ

$$\lim_{n\to\infty}(1 - F(v_n)) = 1,\ 1 - \gamma^{-1} \text{ または } 1 + \gamma^{-1}$$

が成り立つ．また，上で定義した $\{a_n\}$，$\{b_n\}$ に対しては

$$\lim_{n\to\infty}(v_n - b_n)/a_n = 0,\ (1 - \gamma^{-1})^{-(1/\gamma)} \text{ または } \quad (-1)(1 + \gamma^{-1})^{(1/\gamma)}$$

および

$$\lim_{n\to\infty} n(1 - F(v_n)) = \lim_{n\to\infty} \frac{P(\xi_1 > v_n)}{P(M_n > v_n)}$$

が成り立つ．

なお，$F \in \mathsf{D}(G_{-\gamma})$ のための必要十分条件は $1 - F \in RV_{-\gamma}$ である(Niu (1997)参照)．

2.2 Poisson-like 確率変数の最大値の極限分布

独立な Poisson 確率変数の最大値の分布はどのような正規化数列を選んでも非退化極限分布に収束しないことは知られている．一方，平均値が大きければ Poisson 分布は正規分布で近似することができ，しかも正規確率変数列の最大値の分布は $G_0(x)$ で近似することができる．それでは「どのような Poisson 確率変数列の最大値ならばその分布を $G_0(x)$ で近似できるか」という問題が，1986 年のチェルノブイリ原子力発電所の事故に関連して，γ-線をモニターする必要から起こった．

各自然数 n に対して $R_{n,i}$ $(i = 1, \cdots, n)$ を平均値 λ_n をもつ i.i.d. Poisson 確率変数列とする．次に，X を原点の近くで積率母関数をもち，$EX = 0, EX^2 = \sigma^2$ の確率変数とする．$X_1, X_2, \cdots$ を X の独立なコピーとし

$$S_n = \sum_{i=1}^{n} X_i$$

とおく．そのとき，Cramér の定理により，$x \to \infty$，$x = o(\sqrt{n})$ に対し

$$\frac{P(S_n/(\sigma\sqrt{n}) > x)}{1 - \Phi(x)} = \exp\left(x^2 C\left(\frac{x}{\sqrt{n}}\right)\right)\left(1 + O\left(\frac{x}{\sqrt{n}}\right)\right)$$

とかくことができる．ただし，$\Phi(x)$ は標準正規分布関数であり，級数

$$C(z) = c_1 z + c_2 z^2 + \cdots$$

の各 j に対する係数は EX_1^i $(1 \le i \le j+2)$ の関数として表されるものである（Petrov（1975）参照）．

$$h_{n,r}(x) = \frac{x^2}{2} + \log x + \frac{1}{2}\log 2\pi - x^2 \sum_{j=1}^{r} c_j \left(\frac{x}{\sqrt{\lambda_n}}\right)^j$$

とおく．$b_n^{(r)}$ $(r \ge 1)$ を

$$h_{n,r}(x) = \log n$$

の $x = O(\sqrt{\lambda_n})$ の範囲での解とすると，解は唯 1 つで

$$b_n^{(r)} \sim (2\log n)^{\frac{1}{2}} \quad (n \to \infty)$$

となり

$$a_n = \frac{1}{h_n'(b_n^{(r)})}$$

とおくと，$a_n \sim (2\log n)^{-\frac{1}{2}}$ となる．特に，$r = 0$ のときは

$$a_n = (2\log n)^{-\frac{1}{2}}$$
$$b_n^{(0)} = (2\log n)^{\frac{1}{2}} - \frac{\log\log n + \log 4\pi}{2(2\log n)^{\frac{1}{2}}}$$

となる．Anderson et al.（1997）は次の定理 2.2，2.3 を証明した．

定理 2.2. ある整数 r (≥ 0) に対して λ_n が

$$\log n = o\left(\lambda_n^{\frac{r+1}{r+3}}\right) \quad (n \to \infty)$$

を満たしながら増加するとき，

$$u_n(x) = \lambda_n + \sqrt{\lambda_n}(b_n^{(r)} + a_n x)$$

とおけば

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} R_{n,i} \le u_n(x)\right) = G_0(x)$$

が成り立つ．

$\lambda_n = o(n)$ とすると，次の定理が成り立つ．

定理 2.3. $R_{n,i}$ $(1 \le i \le n)$ を平均値 λ_n の i.i.d. Poisson 確率変数列とする．そのとき

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} R_{n,i} = J_n \text{ or } J_n + 1\right) = 1$$

を満たす整数列 $\{J_n\}$ が存在する．

定理 2.3 から，$\lambda_n = o(\log n)$ の場合は，λ_n が定数の場合と同様になることがわかる．

3. 強混合型定常過程の極値の性質

3.1 順序統計量

$\{\xi_n; n \ge 1\}$ を強混合型定常確率変数列とする．順序統計量 $\xi_{n:j}$ を用いて $\eta_{n,i}$ を次のように定義する：

$$\eta_{n,i} = \xi_{n:n-i+1},\ (i < n); \quad = \xi_{n:1}(i \ge n)$$

したがって，$M_n = \eta_{n,1}$ である．

Welsch (1972) は次の定理を証明した．

定理 3.1. ξ_1 の分布関数 F はある正規化数列 $\{a_n\}$，$\{b_n\}$ と極限分布関数 $G(x)$ に対して

$$F^n(a_n x + b_n) \to G(x)$$

を満たすものとする．さらに

$$k_n \to \infty,\ p_n \to \infty,\ \frac{n}{k_n p_n} \to 1,\ k_n^2 \alpha\left(\left[\frac{n - k_n p_n}{k_n}\right]\right) \to 0$$

$$\lim_{n\to\infty} \sum_{j=1}^{p_n - 1} (p_n - j) P(\xi_1 > a_n x + b_n, \xi_{j+1} > a_n x + b_n) = 0 \quad {}^\forall x > 0$$

を満たす整数列 $\{k_n\}$，$\{p_n\}$ が存在すれば，すべての x, y に対して

$$P(\eta_{n,1} > a_n x + b_n, \eta_{n,2} > a_n y + b_n) \to H(x, y) \quad {}^\forall x, y$$

が成り立つ．ただし，

$$H(x, y) = G(y)\left(1 + \log\frac{G(x)}{G(y)}\right)\ (y < x); \quad = G(x)\ (y \ge x)$$

とする．

次に

$$m_n(t) = \eta_{[nt],1}, \quad s_n(t) = \eta_{[nt],2} \quad (0 \le t \le 1)$$

とおく．

定理 3.2. 定理 3.1 と同じ条件の下で

$$\{(m_n(t), s_n(t)) : 0 \le t \le 1\} \xrightarrow{D} \{(m(t), s(t)) : 0 \le t \le 1\} \quad \text{in } \mathcal{D}^2[0, 1]$$

が成り立つ．ただし $(m(t), s(t))$ の分布は

$$
\begin{aligned}
&P(m(t_1) \le x_1, s(t_1) \le y_1; m(t_2) \le x_2, s(t_2) \le y_2) \\
&= G^{t_1}(y_1)\left(1 + t_1 \log \frac{G(x_1)}{G(y_1)}\right) G^{t_2-t_1}(y_2)\left(1 + (t_2 - t_1)\log \frac{G(x_2)}{G(y_2)}\right) \\
&\qquad (0 < t_1 \le t_2, y_1 \le x_1 \le y_2 \le x_2) \\
&= G^{t_1}(y_1) G^{t_2-t_1}(y_2)\left(1 + t_1 \log \frac{G(x_1)}{G(y_1)}\right) \\
&\quad + G^{t_1}(y_1)\left(1 + t_1 \log \frac{G(y_2)}{G(y_1)}\right) G^{t_2-t_1}(y_2)\left(1 + (t_2 - t_1)\log \frac{G(x_2)}{G(y_2)}\right) \\
&\qquad (y_1 \le y_2 \le x_1 \le x_2)
\end{aligned}
$$

である．

高次元の場合も同様に得られる．

3.2 標本平均と稀有事象

$\{\xi_i\}$ を強混合型定常確率変数列で，$E\xi_1 = 0$，$E\xi_1^2 < \infty$ とし

$$
S_n = \sum_{i=1}^{n} \xi_i, \quad \sigma_n^2 = ES_n^2
$$

とおく．

Hsing（1995a，1995b）は次の定理を証明した．

定理 3.3. （$\sigma_n^2 \to \infty$ で）$\{S_n/\sigma_n\}$ に対して中心極限定理が成り立つものとし，$\{B_n\}$ は R の Borel 集合の任意の系列で

$$
\liminf_{n\to\infty} P\left(\bigcap_{i=1}^{n} \{\xi_i \in B_n\}\right) > 0
$$

を満たすものとする．そのとき

$$
P\left(\frac{S_n}{\sigma_n} \le x \Big| \bigcap_{i=1}^{n} \{\xi_i \in B_n\}\right) \xrightarrow{D} \Phi(x)
$$

が成り立つ．

定理 3.3 から，ある単調増加関数 $u_n(y)$ $(y \in \mathsf{R})$ に対して

$$
P(M_n \le u_n(y)) \xrightarrow{D} G(y)
$$

(G は極限分布関数) ならば

$$
\lim_{n\to\infty} P\left(\frac{S_n}{\sigma_n} \le x, M_n \le u_n(y)\right) = \Phi(x)G(y)
$$

が成り立つことがわかる．

3.3 応用–治癒過程

動植物の細胞や器官は破壊されると，時間をかけてその破壊された部位を治そうとする．Hüsler（1995）はこの再生治癒に着目して次の二つのモデルを考えた．

モデル 1．　破壊された部位に関係のない再生の問題

A_i $(i \leq n)$ は i 番目に破壊された部位を表し，再生は D_i 時間遅れて早さ p_n で開始されるとする．そのとき，部位 i 個々の再生に要する時間 τ_i は

$$\tau_i = \frac{A_i}{p_n} + D_i$$

で表される．ただし，$\{A_i\}$，$\{D_i\}$ はそれぞれ i.i.d. で，互いに独立とする．この場合，個々の部位の再生ではなく，全体の再生に要する時間が問題となるので

$$\tau = \max_{1 \leq i \leq n} \tau_i$$

を考える．

(i) D_i が定数の場合

$\{A_i^*\}$ を i.i.d. で A_1^* の分布を F_{A^*}，$A_i = q_n A_i^*$ とし, $d_n > 0$ を定数とする．そのとき

$$\tau = \frac{q_n}{p_n} \max_{1 \leq i \leq n} A_i^* + d_n$$

となるので，次の定理が成り立つ．

定理 3.4.　$D_i = d_n$ a.s. が成り立つとき，k $(= 0, -r, r)$ に対して $F_{A^*} \in \mathsf{D}(G_k)$ ならば

$$P\left(\tau \leq \frac{q_n}{p_n}(a_n x + b_n) + d_n\right) \xrightarrow{D} G_k(x)$$

が成り立ち，また，逆も成り立つ．

(注)この結果は $\{A_i\}$ が弱従属条件を満たす定常列の場合にも成り立つ(Leadbetter et al. (1983)参照)．

(ii) D_i が確率変数で，A_i が定数の場合

D_i の分布を F_D とする．そのとき

$$\tau = \frac{q_n}{p_n} + \max_{1 \leq i \leq n} D_i$$

となるので，次の定理が成り立つ．

定理 3.5.　$A_i = q_n$ a.s. が成り立つとき，k $(= 0, -r, r)$ に対して $F_D \in \mathsf{D}(G_k)$ となれば

$$P\left(\tau \leq \frac{q_n}{p_n} + a_n x + b_n\right) \xrightarrow{D} G_k(x).$$

モデル 2．破壊された部位が関係する場合

端の影響を考えるため，部位が線上あるいは円周上にある場合を考える．切り傷や道路の窪みがその例である．与えられた部位が半径 1 の円の周であり，その周上の n 個所がランダムに破壊されたとする．各個の右側の端の点を ξ_i，弧の長さを A_i とする．A_i も ξ_i も確率変数と考え，ξ_i は円周上で一様分布をしていると仮定する．このとき，$\{A_n; n \geq 1\}$ と $\{\xi_n; n \geq 1\}$ は互いに独立であるとし，D_i 時間遅れて再生が始まると考え，個々の再生時間を

$$\tau_i = \frac{1}{p_n} \min(R_i, A_i) + D_i \tag{3.1}$$

とおく．ただし，$R_i (i \le n)$ は n 個の点 $\xi_1, \cdots, \xi_n$ の間隔(spacing)を表す．そのとき

$$\tau = \max_{1 \le i \le n} \left\{ \frac{1}{p_n} \min(R_i, A_i) + D_i \right\}$$

と定義する．

(3.1)によって定義された $\{\tau_i\}$ は独立ではないが，間隔の取り方の表現から，多くの場合は

$$\tau^* = \max_{1 \le i \le n} \left\{ \frac{1}{p_n} \min\left(\frac{Y_i}{n}, A_i \right) + D_i \right\}$$

と表される．ここで，$\{Y_i\}$ は i.i.d. で，Y_1 はパラメータ 1 の指数分布に従い，$\{Y_i\}$ は $\{A_i\}$, $\{D_i\}$ と独立とする．そのとき，次の定理が成り立つ．

定理 3.6.

$$P(\tau^* \le u_n(x)) \xrightarrow{D} G$$

と

$$P\left(\max_{1 \le i \le n} \left\{ \frac{1}{p_n} \min\left(\frac{Y_i}{n}\left(1 \pm \frac{K}{\sqrt{n}}\right), A_i \right) + D_i \right\} \le u_n(x) \right) \xrightarrow{D} G \quad {}^{\forall} K > 0$$

とは同値である．

4. 極値指標の推定

4.1 独立標本の場合

$\{\xi_n; n \ge 1\}$ は i.i.d. とし，その共通の分布関数を F とする．そのとき，適当な 正規化数列 $\{a_n\}$，$\{b_n\}$ を選べば，極限分布の表現を変えると

$$\mathcal{L}(a_n^{-1}(M_n - b_n)) \xrightarrow{D} G_\gamma^* \tag{4.1}$$

が成り立つ．ただし，

$$\begin{aligned} G_\gamma^*(x) &= \exp(-(1+\gamma x)^{-\frac{1}{\gamma}}) \quad (1+\gamma x > 0, \gamma \ne 0) \\ &= G_0(x) \quad (\gamma = 0) \end{aligned}$$

とする．

ここで

$$G_0^*(x) = G_0(x) = \lim_{\gamma \to 0} G_\gamma^*(x)$$

が成り立つことに注意する．この γ を極値指標という．(4.1)は

$$n(1 - F(a_n x + b_n)) \to -\log G_\gamma^*(x) \quad (x \in \mathsf{R})$$

と同値である．このとき，$F \in \mathsf{D}(G_\gamma^*)$ とかく．

γ の推定量としてよく使われるものに，Pickands (1975)による推定量

$$\hat{\gamma}_n(i) = \frac{1}{\log 2} \log \frac{\xi_{n:n-i+1} - \xi_{n:n-2i+1}}{\xi_{n:n-2i+1} - \xi_{n:n-4i+1}}$$

と Hill (1975)による推定量

$$H_{n:k} = \frac{1}{k}\sum_{i=1}^{k}\log\frac{\xi_{n:n-i+1}}{\xi_{n:n-k}} \tag{4.2}$$

がある．強混合型条件を満たす標本を考えるときは，前者より後者の方が扱いやすい．

$\{\xi_n\}$ が Pareto 分布，すなわち

$$1 - F(x) = x^{-\alpha} \quad (x > 1)$$

に従う場合，$\log\xi_1, \cdots, \log\xi_n$ の分布はパラメータ α の指数分布に従うので，それらの平均は α^{-1}，その最尤推定量は標本平均

$$H_n = \frac{1}{n}\sum_{i=1}^{n}\xi_i = \frac{1}{n}\sum_{i=1}^{n}\xi_{n:n-i+1}$$

となる．もし Pareto 分布の代わりに一般 Pareto 分布

$$1 - F(x) = x^{-\alpha}L(x)\ (x \to \infty) \quad (L(x)\ \text{は緩変化関数})$$

ならば，Hill 推定量(4.2)を考えると便利である．これは，大雑把にいうと，Pareto 分布に似た分布をもつ場合，大きい方から $k+1$ 番目までのデータを推定に利用しようとするものである．Hill 推定量の性質については詳しく調べられている．

最初に，$\{\xi_n\}$ を正で非有界な i.i.d. 確率変数列とし，$\{k_n\} \in \mathsf{K}$ に対し

$$W_n = \sum_{i=1}^{k_n}\xi_{n:n-i+1} \tag{4.3}$$

について考える．

Mason (1982)は次の定理を証明した．

定理 4.1. (I) ある正の数 c とすべての数列 $\{k_n\}$ $(k_n = [n^\kappa](0 < \kappa < 1))$ に対して

$$T_n = \frac{1}{k_n}\sum_{i=1}^{k_n}\xi_{n:n-i+1} - \xi_{n:n-k_n} \overset{\text{a.s.}}{\to} c\ (n \to \infty)$$

が成り立つための必要十分条件は

$$F \in \mathcal{E}_c = \left\{F : \lim_{x\to\infty}\frac{1}{1-F(x)}\int_x^\infty (1 - F(y))dy = c\right\}$$

が成り立つことである．

(II) ある正の数 c とすべての数列 $\{k_n\}$ $(k_n = [n^\kappa]\ (0 < \kappa < 1))$ に対して

$$S_n = \frac{1}{k_n}\sum_{i=1}^{k_n}\log\xi_{n:n-i+1} - \log\xi_{n:n-k_n} \overset{\text{a.s.}}{\to} c\ (n \to \infty)$$

が成り立つための必要十分条件は

$$F \in \mathcal{R}_c = \left\{F : {}^\forall t > 0\ \lim_{x\to\infty}\frac{1-F(tx)}{1-F(x)} = t^{-\frac{1}{c}}\right\}$$

が成り立つことである．

(注) 定理 4.1 (II)は Hill 推定量と極値指標の関係を示すものである．

次に，$\{k_n\} \in \mathsf{K}$ **で，特に**

$$\limsup_{p\to\infty} \frac{k_{n_{p+1}}}{k_{n_p}} = 1 \quad \left(n_p = \left[\exp\left(\frac{p}{\log p}\right)\right]\right)$$

を満たす数列の集合を K^* **とする．また，**

$$\mathcal{E}_c^* = \left\{F : \lim_{x\to\infty} \frac{-\log(1-F(x))}{x} = c\right\}$$

とおく．

Broniatowski and Weber **(**1997**)は次の定理を証明した．**

定理 4.2. $\{\xi_n\}$ **を非負** i.i.d. **確率変数列とし，**ξ_1 **の分布を** F **とする．**(I)$F \in \mathcal{E}_c^*$ **のとき，**$\{k_n\} \in \mathsf{K}^*$ **がさらに**

$$\forall \epsilon > 0 \quad k_n = o\left(n(\log n)^{-\epsilon}\right) \ (n\to\infty) \tag{4.4}$$

を満たせば

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^{k_n} \xi_{n:n-i+1}}{k_n \log(n/k_n)} = \frac{1}{c} \quad \text{a.s.}$$

が成り立つ．

(II**)** $\{k_n\} \in \mathsf{K}^*$ **が**

$$\liminf_{n\to\infty} \frac{k_n}{\log\log n} \geq 1 \tag{4.5}$$

を満たし，F **が**

$$\liminf_{n\to\infty} \frac{-\log(1-F(x))}{x} \leq c$$

を満たせば

$$\liminf_{n\to\infty} \frac{\sum_{i=1}^{k_n} \xi_{n:n-i+1}}{k_n \log(n/k_n)} \geq \frac{1}{c} \quad \text{a.s.}$$

が成り立つ．

(III**) さらに，**$F \in \mathcal{E}_c^*$ **ならば(**4.4**)，(**4.5**)を満たすすべての** $\{k_n\} \in \mathsf{K}^*$ **に対して**

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{k_n} \xi_{n:n-i+1}}{k_n \log(n/k_n)} = \frac{1}{c} \quad \text{a.s.}$$

が成り立つ．

(注) $F \in \mathcal{E}_c$ **ならば** $F \in \mathcal{E}_{(1/c)}^*$ **である．また，**$F \in \mathcal{E}_{(1/c)}^*$**，すなわち**

$$-\log(1-F(x)) = \frac{x}{c}(1+\epsilon(x))$$

ならば

$$\lim_{x\to\infty} \epsilon(x) = 0 \quad \to \quad F \in \mathsf{D}(G_0^*)$$

$$\lim_{x\to\infty} x\epsilon(x) = 0 \quad \to \quad F \in \mathcal{E}_c$$

が成り立つ．

4.2　定常列に対する Hill 推定量の一致性

定常列に対する Hill 推定量の一致性については，次の結果が得られている．$\{\xi_n\}$ を定常な確率変数列で，その周辺分布 F が

$$1 - F(x) = x^{-\alpha}L(x)$$

を満たすものとする．$\{\xi_n\}$ が下記の何れかの場合，$\{k_n\} \in \mathsf{K}$ ならば

$$H_{n:k} \xrightarrow{D} \alpha^{-1}$$

が成り立つ．

(i) $\{\xi_n\}$ が i.i.d. の場合(Mason (1982)再掲)
(ii) $\{\xi_n\}$ が弱従属の場合(Hsing (1991))
(iii) $\{\eta_t\}$ が $MA(\infty)$ 過程の場合，すなわち

$$\eta_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \quad (\{\xi_n\}\text{i.i.d.})$$

(Resnick (1997))．

また，i.i.d. の場合には

$$\sqrt{k}(H_{n:k} - \alpha^{-1}) \xrightarrow{D} N(0, \alpha^{-2})$$

が得られている．

しかし，$\{\xi_n\}$ に何らかの従属性があると，このことは必ずしも成り立たない．

4.3　絶対正則条件を満たすデータからつくられた Hill 推定量

確率変数列 $\{\xi_n; n \ge 1\}$ に対して $\mathbf{M}_a^b = \sigma(\xi_i : a \le i \le b)$ とおく．$\{\xi_n\}$ が

$$\beta(k) = \sup_{\ell\ge1} E\left(\sup_{A\in\mathbf{M}_{\ell+k+1}^{\infty}} |P(A|\mathbf{M}_1^{\ell}) - P(A)|\right) \to 0 \quad (k\to\infty)$$

を満たすとき，$\{\xi_n\}$ は絶対正則条件または β-混合型条件を満たすという．

$$\beta(k) \le \alpha(k)$$

が成り立つので，絶対正則ならば強混合型条件を満たすが，逆は必ずしも成り立たない．時系列などでその例をつくることは可能である．絶対正則な時系列の例としては，ARMA-，ARCH-，GARCH-モデルを挙げることができ，また，その中に，$k\to\infty$ のとき $\beta(k) = O(e^{-\lambda k})(\lambda > 0)$ を満たすものも数多く存在することが知られている(Yoshihara (1992，2001)参照)．

ここで，$\{\xi_n\}$ は定常絶対正則な確率変数列で，その周辺分布 は F とする．

$F \in \mathsf{D}(G_\gamma^*)$ であるための必要十分条件は

$$R(\lambda : t) = \frac{F^{-1}(1-\lambda t) - F^{-1}(1-\lambda)}{a(\lambda)} - F_\gamma^{-1}(t) \to 0 \quad (\lambda \downarrow 0)$$

が成り立つことである．ただし，$a : (0,1) \to (0,\infty)$ は正規化関数で

$$F_\gamma^{-1}(t) = (-\log G_\gamma)^{-1}(t) = \frac{t^{-\gamma}-1}{\gamma} \quad (\gamma \ne 0)$$

$$= -\log t \quad (\gamma = 0)$$

とする．このとき，荷重関数 $q\ (\in \mathcal{D}[0,1])$ が条件

$$^{\exists}\nu \in \left[0, \frac{1}{2}\right]\ ^{\exists}\mu(>0) \qquad \inf_{x\in[\theta,1]} q(x) > 0\ ^{\forall}\theta > 0; \quad x^{\nu}|\log x|^{\mu} = O(q(x))\ (x \downarrow 0)$$

を満たせば，$t_0 > 1$ に対して

$$k_n^{\frac{1}{2}} \sup_{t\in[0,t_0]} \frac{t^{\gamma+1}}{q(t)} \left| R\left(\frac{k_n}{n}, t\right)\right| \to 0 \tag{4.6}$$

を満たす整数列 $\{k_n\} \in \mathsf{K}$ が存在することが知られている．

ここで，$F \in \mathcal{D}(G_{\gamma}^{*})$ の場合を考える．F_n を $\{\xi_n\}$ からつくられた経験分布，F_n^{-1} をその逆関数とし，(4.6)を満たす $\{k_n\}$ に対し尾部経験分位関数

$$Q_n(t) = F_n^{-1}\left(1 - \frac{k_n}{n}t\right) = \xi_{n:n-[k_n t]} \quad (t \in [0,1])$$

を定義する．さらに，整数列 $\{\ell_n\}$ は次の 3 条件を満たすものとする．

C1．
$$\lim_{n\to\infty} \max\left\{\frac{\beta(\ell_n)}{\ell_n}n, \frac{\ell_n}{k_n^{\frac{1}{2}}}\log^2 k_n\right\} = 0;$$

C2．すべての $0 \le x, y \le 1$ に対して極限

$$\lim_{n\to\infty} \frac{n}{\ell_n v_n}\mathrm{cov}\left(\sum_{i=1}^{\ell_n} I\left(\xi_i > F^{-1}\left(1 - \frac{k_n}{n}x\right)\right), \sum_{i=1}^{\ell_n} I\left(\xi_i > F^{-1}\left(1 - \frac{k_n}{n}y\right)\right)\right)$$

が存在する．この極限を $r(x,y)$ とかく．

C3．すべての $0 \le x, y \le 1+\epsilon$ とすべての $n \ge 1$ に対して

$$\frac{n}{\ell_n v_n}\mathrm{Var}\left(\sum_{i=1}^{\ell_n} I\left(F^{-1}\left(1 - \frac{k_n}{n}y\right) < \xi_i < F^{-1}\left(1 - \frac{k_n}{n}x\right)\right)\right) \le C(y-x)$$

を満たす 定数 C が存在する．ただし，$0 < \epsilon < 1$ とする．

そのとき，q に少し条件をつけると次の定理が成り立つ．

定理 4.3. $\{\xi_n : n \ge 1\}$ が連続な周辺分布 $F \in \mathsf{D}(G_{\gamma}^{*})$ をもつとき，上の条件が成り立てば

$$\sup_{t\in(0,1]} \frac{t^{\gamma+1}}{q(t)} \left| k_n^{\frac{1}{2}}\left(\frac{Q_n(t) - D_n}{a(k_n/n)} - F_{\gamma}^{-1}(t)\right) - t^{-\gamma-1}\Gamma(t)\right| \xrightarrow{P} 0$$

を満たす Q_n のヴァージョンと，共分散 r をもつ中心化された Gauss 過程 $\{\Gamma(t) : 0 \le t \le 1\}$ が存在する．ただし，

$$D_n = F^{-1}\left(1 - \frac{k_n}{n}\right) \quad \left(\lim_{t\downarrow 0} \frac{t^{\gamma+1}}{q(t)} = 0\right)$$
$$= \xi_{n:n} + \frac{1}{\gamma}a\left(\frac{k_n}{n}\right) \quad \text{（その他）}$$

とする．

次に，関数空間

$$\mathcal{D}_{\gamma,q} = \left\{ z : [0,1] \to \mathsf{R} \Big| \lim_{t\downarrow 0} \frac{|z(t)|t^{\gamma+1}}{q(t)} = 0,\ \left\{ \frac{z(t)t^{\gamma+1}}{q(t)} : \theta \le t \le 1 \right\} \in \mathcal{D}[0,1] \right\}$$
$$\mathcal{C}_{\gamma,q} = \left\{ z : [0,1] \to \mathsf{R} \Big| \lim_{t\downarrow 0} \frac{|z(t)|t^{\gamma+1}}{q(t)} = 0,\ \left\{ \frac{z(t)t^{\gamma+1}}{q(t)} : \theta \le t \le 1 \right\} \in \mathcal{C}[0,1] \right\}$$

とおく．$\mathcal{D}_{\gamma,q}$ が準ノルム

$$\|z\|_{\gamma,q} = \sup_{t\in(0,1]} \frac{|z(t)|t^{\gamma+1}}{q(t)}$$

をもつものとする．

一般には，F_γ^{-1} は必ずしも $\mathcal{D}_{\gamma,q}$ に属さないので，代わりに

$$\begin{aligned}\bar{F}_\gamma^{-1}(t) &= \frac{1}{\gamma t^\gamma} \quad (\gamma \ne 0)\\ &= -\log t \quad (\gamma = 0)\end{aligned}$$

を考える．さらに

$$\bar{D}_n = D_n - \frac{1}{\gamma} a\left(\frac{k_n}{n}\right) I(\gamma \ne 0) \quad \left(\frac{0}{0} = 0\right)$$

とおくと，定理 4.3 から

$$k_n^{\frac{1}{2}} \left(\frac{Q_n - \bar{D}_n}{a(k_n/n)} - \bar{F}_\gamma^{-1}(t) \right) \xrightarrow{D} \{t^{-\gamma-1}\Gamma(t) : 0 \le t \le 1\} \quad (\text{in } \mathcal{D}_{\gamma,q})$$

が成り立つ．

次に，極値指標の推定量を表す尾部汎関数 $T(Q_n)$ を考える．T は下の条件を満足するものと仮定する：

$T1$. $\quad T(az+b) = T(z) \quad {}^\forall a > 0, b \in \mathsf{R}, z \in \mathcal{D}_{\gamma,q}$

$T2$. $\quad T(\bar{F}_\gamma^{-1}) = \gamma$

$T3$. $\quad {}^\forall \lambda_n\ (\downarrow 0)\ {}^\forall z_n (\in \mathcal{D}_{\gamma,q}) \ \to z (\in \mathcal{C}_{\gamma,q})$

$$\frac{1}{\lambda_n}\{T(\bar{F}_\gamma^{-1} + \lambda_n z_n) - T(\bar{F}_\gamma^{-1})\} \to T'_\gamma(z)$$

ただし，$T'_\gamma : \mathcal{C}_{\gamma,q} \to \mathsf{R}$ は連続な線形汎関数とする．さらに，

$$\int_{[0,1]} t^{-\gamma-1} q(t) |\nu_{T,\gamma}|(dt) < \infty,$$
$$T'_\gamma(z) = \int_{[0,1]} z(t) \nu_{T,\gamma}(dt) \quad {}^\forall z \in \mathcal{C}_{\gamma,q}$$

を満たす $[0,1]$ の上の符号つき測度 $\nu_{T,\gamma}$ がただ一つ存在すると仮定する．

そのとき，(4.6)から $T(Q_n)$ の漸近的正規性が得られる．すなわち

$$\mathcal{L}(k_n^{\frac{1}{2}}(T(Q_n) - \gamma)) \xrightarrow{D} N(0, \sigma_{T,\gamma}^2)$$

ただし，

$$\sigma_{T,\gamma}^2 = \int_{[0,1]^2} (st)^{-\gamma+1} r(s,t) \nu_{T,\gamma}^2(ds, dt).$$

例として，一般 Pareto モデルのパラメータの最尤推定量，すなわち，尤度方程式

$$\frac{1}{k_n}\sum_{i=1}^{k_n}\left(1-\frac{\gamma}{\sigma}(\xi_{n:n-i+1}-\xi_{n:n-k_n})\right)^{-1}=\frac{1}{\gamma+1}$$
$$\frac{1}{k_n}\sum_{i=1}^{k_n}\left(1-\frac{\gamma}{\sigma}(\xi_{n:n-i+1}-\xi_{n:n-k_n})\right)=\gamma$$

$(\gamma>-(1/2))$ の解 $(\hat{\gamma},\hat{\sigma})$ や Hill 推定量 $H_{n:k}$ を挙げることができる．

Resnick and Stărică (1995)は，$1-F\in RV_{-\alpha}$ $(\alpha>0)$ とし，

$$b(t)=F^{-1}\left(1-\frac{1}{t}\right)\quad(t>1)$$

とおき，任意の数列 $\{k_n\}$ に対して $\tilde{\xi}_{(k)}$ を $b(n/k_n)$ の推定量と定義して一致性を証明し，MA(∞) に関して，Hill 推定量の一致性を証明した．

4.4 Hill プロット

$\{\xi_n\}$ を定常確率系列で周辺分布 F は

$$1-F(x)\sim x^{-\gamma}L(x)\quad(L(x)\text{ は緩変化関数})$$

を満たすものとする．$H_{n:k}$ を Hill 推定量とするとき $\{(k,H_{n:k}):1\le k\le n-1\}$ を Hill プロットと呼ぶ．

$$H_{n:k}\xrightarrow{P}\gamma\quad(n\to\infty)$$

であるから，実用的には，Hill プロットをグラフ上にかき込み，グラフが安定する点から γ を推測することができる．i.i.d. の場合

$$b(t)=F^{-1}\left(1-\frac{1}{t}\right)$$

とし，$A:(0,\infty)\to\mathsf{R}$ を，$t\to\infty$ のとき $A(t)$ が定符号となる関数で，ある定数 $\rho<0$ に対し

$$\lim_{t\to\infty}\frac{b(tx)-\log b(t)-\gamma\log x}{A(t)}=\frac{x^\rho-1}{\rho}\quad(x>0)$$

が成り立つものとする．Drees et al. (2000)は次の定理を証明した．

定理 4.4. (4.6)が成り立てば，$H_{n:i}$ $(1\le i\le n-1)$ のヴァージョンと Wiener 過程 W が存在し，$\{j_n\}\in\mathsf{K}$ および $\{\ell_n\}\in\mathsf{K}$ に対して

$$H_{n:i}-\left(\gamma+\gamma\frac{W(i)}{i}+\frac{A(n/i)}{1-\rho}\right)=O\left(\frac{\log i}{i}\right)+o\left(A\left(\frac{n}{i}\right)\right)\quad\text{a.s.}$$

が，$n\to\infty$ のとき，$j_n\le i\le\ell_n$ に対して一様に成り立つ．

さらに，$n\to\infty$ のとき

$$H_{n:i}-\gamma\frac{1}{i}\sum_{j=1}^{i}\xi_j^*=O\left(A\left(\frac{n}{i}\right)\right)\quad\text{a.s.}$$

が $1\le i\le\ell_n$ に対して一様に成立するような標準指数分布をもつ i.i.d. 確率変数列 $\{\xi_n^*\}$ が存在する．

系． $\{k_n\} \in \mathsf{K}$ **とする．(4.6)が成り立てば，定理 4.4 で用いた** $H_{n:k}$ **のヴァージョンに対して**

$$\sup_{t_n \le t \le T_n} (\sqrt{t} \wedge t^{\rho-\tau}) \left| H_{n:[k_n,t]} - \left(\gamma + \frac{1}{\sqrt{k_n}}\gamma \frac{W_n(t)}{t} + A\left(\frac{n}{k_n}\right)\frac{t^{-\rho}}{1-\rho}\right)\right|$$
$$= o_P\left(\frac{1}{\sqrt{k_n}} + A\left(\frac{n}{k_n}\right)\right) \qquad {}^{\forall}\tau > 0$$

が $k_n t_n \to \infty$ **を満たすすべての** $t_n \downarrow 0$ **と** $k_n T_n/n \to 0$ **を満たすすべての** $T_n \uparrow \infty$ **に対して成り立つ．**

さらに， $h(t)$ **が**

$$\lim_{t\to\infty} h(t)\left(\frac{1}{t}\log\log\frac{1}{t}\right)^{\frac{1}{2}} = 0$$

を満たすならば

$$\sup_{0 < t \le T_n} (h(t) \wedge t^{\rho-\tau}) \left| H_{n:[k_n t]} - \left(\gamma + \frac{\gamma}{\sqrt{k_n}}\frac{W_n(t)}{t} + A\left(\frac{n}{k_n}\right)\frac{t^{-\rho}}{1-\rho}\right)\right|$$
$$= o_P\left(\frac{1}{\sqrt{k_n}}\right) + A\left(\frac{n}{k_n}\right)$$

が成り立つ．

（注） $h(t)$ **は** $t/h(t)$ **が** Wiener **過程の上界関数族に入れば，上の系は成り立つ．**

Resnick **(**1997**)では，** F **の尾部に対し**

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\gamma}, \quad (x > 0)$$

が成り立つときの γ **を推定するため，** Hill **プロットを含め，いろいろなプロット方法とそれらの性質を詳しく調べている．**

4.5 $F \in \mathsf{D}(G_0^*)$ vs $F \in \mathsf{D}(G_\gamma^*)(\gamma \neq 0)$ **の検定**

$\xi_1, \cdots, \xi_n$ **を共通な分布関数** F **をもつ独立な確率変数列とし** $F \in \mathsf{D}(G_\gamma^*)$ **と仮定する．**
Marohn **(**1998**)は**

$$F \in \mathsf{D}(G_0^*) \quad \text{vs} \quad F \in \mathsf{D}(G_\gamma^*)(\gamma \neq 0)$$

の検定法を考察した． Gumble **分布** G_0^* **の検定については従来から多く研究されてきたが，局所漸近正規性の枠内で考えられるようになったのは最近のことである(**Marohn **(**1994**)を参照)．**

$$\omega(F) = \sup\{t \in \mathsf{R} : F(t) < 1\} \in (-\infty, \infty]$$

とし，一般 Pareto **分布**

$$V_\gamma(x) = 1 + \log G_\gamma^*(x) \quad (x \ge 0)(\gamma \neq 0)$$
$$V_0(x) = \lim_{\gamma\to 0} V_\gamma(x) = 1 - \exp(-x) \quad (x \ge 0)$$

を考える．したがって， $\gamma \neq 0$ **のときは**

$$V_\gamma(x) = 1 - (1+\gamma x)^{-\frac{1}{\gamma}} \quad (x \ge 0, \gamma > 0;\ 0 \le x \le -1/\gamma, \gamma < 0)$$

となる．v_γ を V_γ の密度関数とする．このとき，ある固定された D と ϵ に対して，F が 一般 Pareto 分布の ϵ-近傍にあるための必要十分な条件は「ある γ に対して

$$\omega(F) = \omega(G_\gamma)$$

が成り立ち，ある $x_0(\gamma) = x_0(\gamma, f) < \omega(F)$ に対して，F は区間 $(x_0, \omega(F))$ の上で微分可能で，f を F の密度関数とするとき

$$\left| \frac{f(x)}{v_\gamma(x)} - 1 \right| \leq D(1 - V_\gamma(x))^\epsilon, \quad x \in (x_0(\gamma), \omega(F))$$

が成り立つこと」である．この条件を満たす F の集合を $\mathcal{Q}_{\epsilon,D}(V_\gamma)$ とおく．

次に，$F_\gamma = V_\gamma$ とし，尺度パラメータ σ を使って最初のモデルを

$$F_{\gamma,\sigma}(x) = F_\gamma\left(\frac{x}{\sigma}\right) \quad x \geq \sigma x_0(\gamma)$$

とかく．また，局所分布型対立仮説として

$$\gamma_n(\theta) = \frac{2\theta k_n^{-\frac{1}{2}}}{\log(n/k_n)},$$

局所尺度対立仮説として

$$\sigma_n(\theta, \rho) = \sigma_n(\theta)\left(1 + \frac{\rho k_n^{-\frac{1}{2}}}{\log(n/k_n)}\right)$$

を考える．ただし，

$$\begin{aligned} \sigma_n(\theta) &= \sigma_0 \frac{\gamma_n(\theta)\log(n/k_n)}{\exp(\gamma_n(\theta)\log(n/k_n)) - 1} \\ &= \sigma_0 \frac{2\theta k_n^{-(1/2)}}{\exp(2\theta k_n^{-(1/2)}) - 1}, \end{aligned}$$

$\sigma_0 > 0$ は固定した定数，$\{k_n\} \in \mathsf{K}$ は

$$k_n^{-(1/2)} \log n \to 0$$

を満たすものとする．パラメータ空間を

$$\Theta \times \Xi_n = \mathsf{R} \times \left(-k_n^{1/2} \log \frac{n}{k_n}, \infty\right)$$

とする．このとき，局所モデルとしては統計的実験

$$\mathsf{E}_{n,k_n,\sigma_0} = \left(\mathsf{R}^{k_n}, \mathbf{B}^{k_n}, \left\{\mathcal{L}((\xi_{n:n-k_n+1}, \cdots, \xi_{n:n}) \middle| F^n_{\gamma_n(\theta),\sigma_n(\theta,\rho)}) : (\theta, \rho) \in (\Theta \times \Xi_n)\right\}\right)$$

を使い，検定問題

仮説 $\{(\theta, \rho) : \theta = 0, \rho \in \mathsf{R}\}$, **対立仮説** $\{(\theta, \rho) : \theta \neq 0, \rho \in \mathsf{R}\}$,

を考える．

最初に，$(\mathsf{E}_{n,k_n,\sigma_0})_n$ の局所漸近正規性について考える．

定理 4.5. $F_\gamma = V_\gamma$ $(\gamma \in \mathsf{R})$ と仮定する．そのとき，$(0, \sigma_0)$ を基にした尤度比は

$$\frac{d\mathcal{L}((\xi_{n:n-k_n+1}, \cdots, \xi_{n:n})|F^n_{\gamma_n(\theta),\sigma_n(\theta,\rho)})}{d\mathcal{L}((\xi_{n:n-k_n+1}, \cdots, \xi_{n:n})|F^n_{0,\sigma_0})}$$

$$= \exp\left\{\langle(\theta,\rho)^T, (Z_{1,n,\sigma_0}, Z_{2,n,\sigma_0})^T\rangle - \frac{1}{2}\|(\theta,\rho)^T\|^2 + o_{F^n_{0,\sigma_0}}(1)\right\}$$

とかくことができる．ただし，

$$Z_{1,n,\sigma_0} = \frac{1}{\sqrt{k_n}}\sum_{j=1}^{k_n}\left(\frac{\xi_{n:n-j+1}}{\sigma_0} - \frac{\xi_{n-k_n+1}}{\sigma_0} - 1\right)$$

$$Z_{2,n,\sigma_0} = \sqrt{k_n}\left(\frac{\xi_{n:n-k_n+1}}{\sigma_0} - \log\frac{n}{k_n}\right)$$

とし，$\langle\cdot,\cdot\rangle$ は内積，$\|\cdot\|$ はそれに対応するノルムを表す．

さらに，$\{k_n\}$ が条件

$$k_n\left(\frac{k_n}{n}\right)^{\epsilon} \to 0 \quad (n\to\infty)$$

を満たせば

$$F_\gamma \in \mathcal{Q}_{\epsilon,D}(V_\gamma)$$

が成り立つ．

この定理から次の検定が可能になる．

系．

仮説 $\{\mathcal{L}((\xi_{n:n-k_n+1},\cdots,\xi_{n:n})|F^n_{\gamma_n(\theta),\sigma_n(\theta,\rho)}) : \theta = 0, \rho\in\Xi_n\}$

対立仮説 $\{\mathcal{L}((\xi_{n:n-k_n+1},\cdots,\xi_{n:n})|F^n_{\gamma_n(\theta),\sigma_n(\theta,\rho)}) : \theta \neq 0, \rho\in\Xi_n\}$

の有意水準 α の漸近的最良不偏検定は

$$\phi^*_{n,k} = \begin{cases} 1 & (|Z_{1,n\hat{\sigma}_n}| > u_{1-(\alpha/2)}) \\ 0 & (|Z_{1,n\hat{\sigma}_n}| \le u_{1-(\alpha/2)}) \end{cases}$$

で与えられる．すなわち，検定関数列 $\{\phi^*_{n,k_n}\}$ は検定力関数の上界

$$\lim_{n\to\infty} E_{\gamma_n(\theta),\sigma_n(\theta,\rho)}\phi^*_{n,k_n} = 2 - \Phi(u_{1-(\alpha/2)}+\theta) - \Phi(u_{1-(\alpha/2)}-\theta) \quad (\rho,\theta\in \mathsf{R})$$

に到達する．ただし，$u_\alpha = \Phi^{-1}(\alpha)$ $(0<\alpha<1)$ である．

参 考 文 献

Anderson, C. W., Coles, S. G. and Hüsler, J. (1997). Maxima of Poisson-like variables and related triangular arrays, *The Annals of Applied Probability*, **7**, 953–971.

Broniatowski, M. and Weber, M. (1997). Strong laws for sums of extreme values, *Theory of Probability and Its Applications*, **42**, 395–404.

Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot, *The Annals of Statistics*, **28**, 254–274.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **3**, 1163–1174.

Hsing, T. (1991). On the tail index estimation using dependent data, *The Annals of Statistics*, **19**, 1547–1569.

Hsing, T. (1995a). On the asymptotic independence of the sum and rare values of weakly dependent random variables, *Stochastic Processes and Their Applications*, **60**, 49–64.

Hsing, T. (1995b). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables, *The Annals of Probability*, **23**, 938–947.

Hüsler, J. (1995). A note on a model for restoring a destroyed region, *Journal of Applied Probability*, **32**, 756–767.

Kennedy, D. P. and Kertz, R. P. (1991). The asymptotic behavior of the reward sequence in the optimal stopping I.I.D. random variables, *The Annals of Probability*, **19**, 329–341.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Marohn, F. (1994). On testing the exponential and Gumbel distribution, *Extreme Value Theory* (eds. J. Galambos et al.), 159–174, Kluwer, Dordrecht.

Marohn, F. (1998). An adaptive efficient test for Gumbel domain of attraction, *Scandinavian Journal of Statistics*, **25**, 211–224.

Mason, D. (1982). Laws of large numbers for sums of extreme values, *The Annals of Probability*, **10**, 754–764.

Niu, X.-F. (1997). Extreme value theory for a class of nonstationary time series with applications, *The Annals of Applied Probability*, **7**, 508–522.

Petrov, V. V. (1975). *Sums of Independent Random Variables*, Springer, Berlin.

Pickands, III, J. (1975). Statistical inference using extreme order statistic, *The Annals of Statistics*, **3**, 119–131.

Resnick, S. I. (1997). Heavy tail modelling and teletraffic data, *The Annals of Statistics*, **25**, 1805–1869.

Resnick, S. and Stărică, C. (1995). Consistency of Hill's estimator for dependent data, *Journal of Applied Probability*, **32**, 139–167.

Rosenblatt, R. (1956). A central limit theorem and a strong mixing condition, *Proceedings of National Academy of Sciences, USA*, **42**, 43–47.

Welsch, R. E. (1972). Limit laws for extreme order statistics from strong-mixing processes, *The Annals of Mathematical Statistics*, **43**, 439–446.

Yoshihara, K. (1992). *Summation Theory for Weakly Dependent Sequences*, Weakly Dependent Stochastic Sequences and Their Applications, Vol. I, Sanseido, Tokyo, Japan.

Yoshihara, K. (2001). *Random Sums, Extremes and Sequential Analysis*, Weakly Dependent Stochastic Sequences and Their Applications, Vol. XII, Sanseido, Tokyo, Japan.

Recent Topics on the Extreme Theory Based on Weakly Dependent Data

Ken-ichi Yoshihara

(Soka University)

From both the theoretical and applied view points, the extremal theory contains many very important problems in various fields. Therefore, the theory has been studied by many authors and is discussed extensively now. Almost all results are based on independent sample, but to apply the results to real problems, the assumption that the sample is independent is too restrictive and in some cases the results obtained can not be used.

In this paper, we survey some recent results on the extreme theory based on weakly dependent data and examine "whether the results under independence remain true or not", "whether we can use the results under independence, if modified" and "whether we can obtain new type results".

Key words: Weak dependence, extremal index, Hill estimator, stationary sequence.