数理解析研究所講究録
第197巻 1974年 70-79

極値統計量の漸近分布と漸近速度

香川大・経　松縄　規

§1. 序

確率変数列 $\{X_i\}$ $(i=1,2,\cdots,N)$ からの順序統計量を $X_{N,1}<X_{N,2}<\cdots<X_{N,N}$ とする. $\{X_i\}$ $(i=1,2,\cdots,N)$ が独立に, steadily increasing な一次元連続分布 (pdf. $f(x)$, cdf. $F(x)$) に従うとき, intermediate term $X_{N,k}$ $(k(N)/N\to 0, \text{ as } N\to\infty)$ は漸近極値性と漸近正規正の二性質を持つことを示す。 $F(x)$ が一様分布のとき $X_{N,k}$ が漸近極値性(の漸近正規性)をより顕著に示す $k=k(N)$ の範囲を, $X_{N,k}$ の exact な分布とそれに対応する二つの漸近分布の漸近速度を評価することにより考察する. 漸近極値分布に対しては distribution-free な誤差評価を行なう.

更に $\{X_i\}$ $(i=1,2,\cdots,N)$ が $m(N)$-dependent な確率変数列であるとき, $X_{N,k}$ の適当な規準化変数に対する漸近分布を, かなり一般的立場から考察し, Watson [7] の stationary process からの最大値に対する結果 $(m: fix)$ の拡張が可能なことを示す.

§2. Intermediate term の漸近的挙動

$\{X_i\}$ $(i=1,2,\cdots,N)$ を区间 (a,b), (a,b: extended real), で独立に同一の連続型分布 (pdf. $f(x)$, cdf. $F(x)$) に従う確率変数列とする。これからの size N の順序統計量 $X_{N,1}<X_{N,2}<\cdots<X_{N,N}$ の第 k 番最小値 $X_{N,k}$ に於る rank k が $N\to\infty$ で

$$(2\cdot1)\qquad k/N\to 0\,,\quad k=k(N)\to l \ \text{(extended positive integer)}$$

を満すとき，$X_{N,k}$ を intermediate term と呼ぶことにする。

$X_{N,k}$ の pdf. は，

$$(2\cdot2)\qquad g_N(x)=\frac{N!}{(k-1)!\,(N-k)!}F^{k-1}(x)\{1-F(x)\}^{N-k}\cdot f(x),\quad (a<x<b)$$

であるが，これに対応して pdf. が次式

$$(2\cdot3)\qquad \tilde{g}_N(x)=\frac{N^k}{\Gamma_k(N)\,\Gamma(k)}F^{k-1}(x)\cdot e^{-NF(x)}\cdot f(x),\qquad (a<x<b)$$

で与えられる確率変数 $\tilde{X}_{N,k}$ を考える。こゝに

$$(2\cdot4)\qquad \Gamma_k(N)=\frac{1}{\Gamma(k)}\int_0^N e^{-t}\,t^{k-1}\,dt\,.$$

また $Z_{N,k}$ を pdf. が次式で与えられる正規確率変数とする:

$$(2\cdot5)\qquad h_N(x)=\frac{1}{\sqrt{2\pi}\,\sigma_{N,k}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_{N,k}}{\sigma_{N,k}}\right)^2\right\},\qquad (-\infty<x<\infty)$$

こゝに，

$$(2\cdot6)\qquad \mu_{N,k}=F^{-1}\left(\frac{k}{N+1}\right),\quad \sigma_{N,k}=\left\{\frac{k(N+1-k)}{(N+1)^2(N+2)\,f^2(\mu_{N,k})}\right\}^{1/2}.$$

$X_{N,k}$ と $\widetilde{X}_{N,k}$ に関する K-L 情報量は次のように求まる：

$$(2\cdot7)\quad I(X_{N,k},\widetilde{X}_{N,k})=\sum_{i=1}^{k-1}\ln\left(1-\frac{i}{N}\right)-(N-k)\sum_{i=1}^{k}\frac{1}{N-k+i}$$
$$+\frac{N}{N+1}k+\ln\Gamma_k(N)$$
$$=\frac{3}{4}\left(\frac{k}{N}\right)^2+\ln\Gamma_k(N)+O\left(max\left(\frac{1}{N^2},\frac{k^3}{N^3}\right)\right).$$

一般の $F(x)$ に対して $I(X_{N,k},Z_{N,k})$ を直接求めるのは容易ではない。 こゝでは以下の正則条件の下で漸近主領域 $Q_{N,k}$ に於る情報量の上限を考えることにする。

$$\begin{cases}\text{条件 } 2\cdot1 & (a,b)=\{x:\ f(x)>0\},\\ \text{条件 } 2\cdot2 & f(x)\in \boldsymbol{C}^1_{(a,b)}\ \ (\ (a,b)\text{上で一回連続微分可能}),\\ \text{条件 } 2\cdot3 & \sup\limits_{z\in Q_{N,k}}\ \sup\limits_{z^*\in((z,\mu_{N,k}))}\ max\{|\varphi(z^*)|,\ \phi(z^*,\mu_{N,k})\}\leqq M,\end{cases}$$

こゝに M は N に無関係な定数，また

$$Q_{N,k}\overset{d}{=}\{\,z:\ 0<l_{N,k}-\delta_{N,k}<z<l_{N,k}+\delta'_{N,k}\}$$

$$\text{with}\begin{cases}\delta_{N,k}=\left\{\dfrac{k}{N+2}l_{N,k}(1-l_{N,k})\right\}^{1/2}\\ \delta'_{N,k}=\left\{\dfrac{N-k}{N+2}l_{N,k}(1-l_{N,k})\right\}^{1/2}\end{cases},\quad l_{N,k}=\frac{k}{N+k},$$

z^* は z と $\mu_{N,k}$ の間に存在する z の或る関数であり，

$$\begin{cases}\varphi(z^*)\overset{d}{=}f'(F^{-1}(z^*))/f^2(F^{-1}(z^*)) & (0<z<1)\\ \phi(z^*,\mu_{N,k})\overset{d}{=}f(F^{-1}(\mu_{N,k}))/f(F^{-1}(z^*)). & (0<z^*,\mu_{N,k}<1)\end{cases}$$

である。

(2·1) 式の条件下で，また $F(x)=x$ なる $(0,1)$ 上の一様分

布の場合には（$Z_{N,k}=W_{N,k}$ と表すことにする），

$$P^{W_{N,k}}(Q_{N,k})\to 1\ ,(N\to\infty) \tag{2·8}$$

となる事がChebycheffの不等式によって示される.(cf.[2]). 確率積分逆変換を行なうことにより，$Q_{N,k}$ 上で

$$I^*(X_{N,k},Z_{N,k})\leqq M\frac{1}{\sqrt{N}}+\sqrt{3}M^2\frac{1}{\sqrt{\min(k,N-k)}}+\frac{3}{2}M^4\frac{1}{\min(k,N-k)} \tag{2·9}$$

と評価できる．

特に，$F(x)=x$ の $(0,1)$ 上の一様分布の場合（$X_{N,k}=U_{N,k}$ と表すことにする），$I(X_{N,k},Z_{N,k})$ はexactに計算されて

$$\begin{aligned}I(U_{N,k},W_{N,k})&=\ln\frac{N!}{(k-1)!(N-k)!}+\ln\sqrt{2\pi}+\frac{1}{2}\ln k+\frac{1}{2}\\&+\frac{1}{2}\ln\left(1-\frac{k+1}{N+2}\right)-(k-1)\sum_{i=1}^{N-k+1}\frac{1}{k-1+i}-(N-k)\sum_{i=1}^{k}\frac{1}{N-k+i},\\&=\frac{1}{3}\left(\frac{1}{k}+\frac{1}{N-k}\right)+O\left(\max\left(\frac{1}{k^2},\frac{1}{(N-k)^2}\right)\right).\end{aligned} \tag{2·10}$$

(2·4) に於て，$1-\Gamma_k(N)\leqq N/(N-k+1)^2$ が成立する事に注意すれば，intermediate term の漸近分布に関して，(2·7)~(2·10) から次の結果を得る．

<u>定理 2·1</u>

(i) $k/N\to 0$ $\Rightarrow X_{N,k}\sim\tilde{X}_{N,k}\ (\mathbb{B})_d,(N\to\infty)$,

(ii) 条件 2·1 ~ 2·3 の下で

$$\begin{cases}(a)\ \min(k,N-k)\to\infty & \Rightarrow X_{N,k}\sim Z_{N,k}\ (\mathbb{B})_d,(N\to\infty)\\(b)\ k/N\to 0,\ k\to\infty & \Rightarrow X_{N,k}\sim\tilde{X}_{N,k}\sim Z_{N,k}\ (\mathbb{B})_d,(N\to\infty),\end{cases}$$

(iii) $$\begin{cases}(a)\ \min(k,N-k)\to\infty & \Rightarrow U_{N,k}\sim W_{N,k}\ (\mathbb{B})_d,(N\to\infty),\\(b)\ k/N\to 0,\ k\to\infty & \Rightarrow U_{N,k}\sim\tilde{U}_{N,k}\sim W_{N,k}\ (\mathbb{B})_d,(N\to\infty),\end{cases}$$

ただし $\tilde{U}_{N,k}$ は (2・3)式で $F(x)=x$ とした pdf. をもつ確率変数である ■

<u>Remark.</u> 上の定理は original な $X_{N,k}$ (or $U_{N,k}$) についての (2・1) の条件下での漸近的挙動についてであるが，適当な<u>規準化</u>を行なった場合の議論が，$k/N \to \infty$, $k \to \infty$ について Chibisov [3], Smirnov [6] によってなされており，もし規準化第 k 番最小値 $X^{*}_{N,k}$ の<u>極限</u>法則が存在するならば，それは<u>正規分布</u>となることが示されており，その為の必十条件も与えられている．■

さて，$X_{N,k} \sim \tilde{X}_{N,k}$ (or $U_{N,k} \sim \tilde{U}_{N,k}$) $(B)_d$ $(N \to \infty)$ が成立するとき $X_{N,k}$ (or $U_{N,k}$) は<u>漸近極値性</u>を持つということにする．(ii)-(b) (or (iii)-(b)) より intermediate term $X_{N,k}$ (or $U_{N,k}$) の漸近分布は $\tilde{X}_{N,k}$ (resp. $\tilde{U}_{N,k}$) と $Z_{N,k}$ (resp. $W_{N,k}$) の mixture :

(2・11) $\quad X_{N,k} \sim \alpha_{N,k}\tilde{X}_{N,k} + (1-\alpha_{N,k})Z_{N,k} \quad (B)_d, \ (N \to \infty),$

or

(2・12) $\quad U_{N,k} \sim \alpha_{N,k}\tilde{U}_{N,k} + (1-\alpha_{N,k})W_{N,k} \quad (B)_d, \ (N \to \infty)$

と考えてもよかろう．こゝに $\alpha_{N,k}$ は $0 \leqq \alpha_{N,k} \leqq 1$ なる定数である．$\alpha_{N,k}$ を具体的に決定することは難しいが，(2・12) の場合については，大凡その検討として，(2・7), (2・10) から

$$\begin{cases} \alpha_{N,k} = 1 & \text{if } k = O(N^{\frac{2}{3}-\varepsilon_1}), \quad 0<\varepsilon_1<2/3, \\ \alpha_{N,k} = 0 & \text{if } k = O(N^{\frac{2}{3}+\varepsilon_2}), \quad 0<\varepsilon_2<1/3, \end{cases}$$

$$0 < \alpha_{N,k} < 1 \quad \text{if} \quad k = O(N^{\delta}), \quad 2/3 - \varepsilon_1 < \delta < 2/3 + \varepsilon_2$$

と見做せる．　(2・11) の場合には情報損失 $I(X_{N,k}, Z_{N,k}) - I^*(X_{N,k}, Z_{N,k})$ を考慮する必要があることや，(2・9) の優評価を用いて上と同様な方法で k の order を定めることには問題があるのでこゝでは扱わない．■

つぎに，$X_{N,k}$ の極値分布 (2・3) への漸近速度について考察する．　一般に確率変数列 $\{X_s\}(s=1,2,\cdots)$，$\{Y_s\}(s=1,2,\cdots)$ に於て，各 s について X_s, Y_s は可測空間 $(R_s, \mathcal{B}_s)$ 上で定義され，共にこの空間上で定義された或る σ-finite measure に関して絶対連続であるとする．　このとき確率測度 P^{X_s}, P^{Y_s} に関して定義された距離：

$$\delta_d(X_s, Y_s) = \sup_{E \in \mathcal{B}_s} |P^{X_s}(E) - P^{Y_s}(E)| \quad (s=1,2,\cdots)$$

と K-L 情報量 $I(X_s, Y_s)$ に関して次が成立する．(証略)

<u>補題 2・1</u>　　$\delta_d(X_s, Y_s) \leqq \sqrt{I(X_s, Y_s)/2} \quad (s=1,2,\cdots)$ ■

この不等式と (2・7) 式より，$X_{N,k}$ の分布と $\tilde{X}_{N,k}$ の分布間の近似誤差に関し次の結果を得る：

<u>定理 2・2</u>

$$\begin{aligned}\delta_d(X_{N,k}, \tilde{X}_{N,k}) &\leqq \Big[\frac{1}{2}\Big\{\sum_{i=1}^{k-1} \ln\Big(1-\frac{i}{N}\Big) - (N-k)\sum_{i=1}^{k}\frac{1}{N-k+i} + \frac{N}{N+1}k + \\ &\quad + \ln\Big(1 - e^{-N}\sum_{i=0}^{k-1}\frac{N^i}{i!}\Big)\Big\}\Big]^{1/2} \qquad (2\cdot 13)\\ &\approx \frac{\sqrt{6}}{4}\frac{k}{N} \quad (\text{for sufficiently large } N)\end{aligned}$$ ■

§3. $m(N)$-dependent sequence の漸近極値分布

Watson [7] は $\{X_i\}$ $(i=1,\cdots,N)$ が m(fix)-dependent で stationary のとき，或る条件下で，最大値 $X_{N,N}$ の極限分布は i.i.d. の場合の極限分布と一致することを示した。 こゝでは Meyer [4], Runnenburg [5] が扱った m-dependent な event sequence の rare events の生起確率を求めることにより，第k番最小値 $X_{N;k}$ の漸近分布を導く。 以下では $m=m(N)$ であり，$\{X_i\}$ $(i=1,\cdots,N)$ は必ずしも stationary でない場合を対象とする。

各Nに対し，$(\Omega_N,\ \mathfrak{a}_N,\ P_N)$ を基礎確率空間，$\{A_i^N\}$ $(i=1,\cdots,N)$ をその中の $m(N)$-dependent な事象列とする。 添数 i の集合 $\{1,2,\cdots,N\}$ の中から次の $r=r(N)$-turples を定義する：

$$
(3\cdot1)\qquad
\begin{cases}
C_r=\{(i_1,\cdots,i_r)\mid 1\leqq i_1<\cdots<i_r\leqq N\}\\
I_r=C_r\cap\{(i_1,\cdots,i_r)\mid i_j-i_{j-1}>m,\ j=2,\cdots,r\}\\
J_r=C_r\cap I_r^c\quad(c:\text{complement})\\
D_r=C_r\cap\{(i_1,\cdots,i_r)\mid i_j-i_{j-1}\leqq m,\ j=2,\cdots,r\}\\
E_r^{(\nu)}=\{(i_1,\cdots,i_r)\mid 1\leqq i_1\leqq i_2\leqq\cdots\leqq i_r\leqq N\ \&\ \text{exactly } \nu \text{ of } i_j\text{'s are equal}\}.
\end{cases}
$$

このとき，Meyer, Runnenburg より一般的な次の結果を得る。

補題3・1

$\{A_i^N\}$ $(i=1,\cdots,N)$ を $m(N)$-dependent な事象列とし，$N\to\infty$ で

$$(3\cdot2)\qquad \sum_{i=1}^{N}P_N(A_i^N)=\xi+o(1),\quad(\xi>0,\ \text{const.}),$$

$$(3\cdot3)\qquad \max_{D_{j+1}}P_N(A_{i_1}^N\mid A_{i_2}^N\cdots A_{i_{j+1}}^N)=O\left(\tfrac{1}{N}\right),\quad(j=1,\cdots,r),$$

(3·4) $\qquad \frac{m}{N}(2\xi)^{r-1} = o(1)$

が満されるならば

(3·5) $\qquad S_r^N \overset{d}{=} \sum_{C_r} P_N(A_{i_1}^N \cdots A_{i_r}^N) = \frac{\xi^r}{r!} + o(1),\ (N \to \infty).$

略証. 次の分割を考える：

(3·6) $\qquad S_r^N = \left(\sum_{I_r} + \sum_{J_r}\right) P_N(A_{i_1}^N \cdots A_{i_r}^N).$

m-dependence に注意して，第1項は

(3·7) $$\sum_{I_r} P_N(A_{i_1}^N \cdots A_{i_r}^N) = \left(\sum_{C_r} - \sum_{J_r}\right)\prod_{j=1}^{r} P_N(A_{i_j}^N)$$

$$= \frac{1}{r!}\left[\left\{\sum_{i=1}^{N} P_N(A_i^N)\right\}^r - \sum_{\nu=2}^{r}\sum_{E_r^{(\nu)}}\prod_{j=1}^{r} P_N(A_{i_j}^N) - \sum_{J_r}\prod_{j=1}^{r} P_N(A_{i_j}^N)\right].$$

条件(3·2)より

(3·8) $\qquad \sum_{E_r^{(\nu)}}\prod_{j=1}^{r} P_N(A_{i_j}^N) = O\left(\frac{1}{N^{\nu-1}}\right),\ (2 \leqq \nu \leqq r),\ (N \to \infty)$

が出る． また(3·3), (3·4)から多少の準備の上，

(3·9) $\qquad \sum_{J_r}\prod_{j=1}^{r} P_N(A_{i_j}^N) = O\left(\frac{m}{N}(2\xi)^{r-1}\right) = o(1),\ (N \to \infty)$

が示せる． 従って(3·6)の第1項は

(3·10) $\qquad \sum_{I_r} P_N(A_{i_1}^N \cdots A_{i_r}^N) = \frac{1}{r!}\xi^r + o(1),\ (N \to \infty).$

また，(3·6)の第2項は(3·9)を示すのと同様にして，条件(3·2)～(3·4)の下で

(3·11) $\qquad \sum_{J_r} P_N(A_{i_1}^N \cdots A_{i_r}^N) = O\left(\frac{m}{N}(2\xi)^{r-1}\right) = o(1)\ \ (N \to \infty)$

となる． よって(3·5)が従う． ∎

さて，こゝで

(3·12) $\qquad P_{[n]}^N \overset{d}{=} P_N(\text{exactly } n \text{ among } \{A_i^N\}\ (i=1,\cdots,N) \text{ occur simultaneously})$

とすると，*inclusion-exclusion formula* により

$$(3\cdot13)\qquad \sum_{t=n}^{n+2l+1}(-1)^{t-n}\binom{t}{n}S_t^N \leqq P_{[n]}^N \leqq \sum_{t=n}^{n+2l}(-1)^{t-n}\binom{t}{n}S_t^N,$$

$$(l=0,1,2,\cdots;\ n+2l+1\leqq N)$$

が成立する．これと補題3·1から次を得る．

<u>補題 3·2</u>　$m(N)$-dependent な事象列 $\{A_i^N\}\ (i=1,\cdots,N)$ が $N\to\infty$ で条件 (3·2)～(3·4) を満すならば

$$(3\cdot14)\qquad P_{[n]}^N \sim \frac{\xi^n e^{-\xi}}{n!}\quad (N\to\infty).\ \blacksquare$$

ところで，いま

$$(3\cdot15)\qquad P_k^N \overset{d}{=} P_N(\text{at least } k \text{ of } \{A_i^N\}\ (i=1,\cdots,N) \text{ occur simultaneously})$$

と置けば

$$(3\cdot16)\qquad P_k^N = 1-\sum_{s=0}^{k-1}P_{[s]}^N$$

であるから，補題3·1により次の結果を得る．

<u>定理 3·1</u>　$m(N)$-dependent な事象列 $\{A_i^N\}\ (i=1,\cdots,N)$ が $N\to\infty$ で条件 (3·2), (3·3) および

$$(3\cdot4)'\qquad \frac{m}{N}(2\xi)^{r-1}\exp\{2\xi r\}=o(1)$$

を満すならば

$$(3\cdot17)\qquad P_k^N \sim \frac{1}{\Gamma(k)}\int_0^{\xi} t^{k-1}e^{-t}dt\quad (N\to\infty)\ \blacksquare$$

<u>系 3·1</u>　$\{X_i\}\ (i=1,\cdots,N)$ を $(\Omega_N, \mathfrak{A}_N, P_N)$ 上の $m(N)$-dependent な実確率変数列，$\{E_N\}=\{(-\infty, C_N(\xi)]\}$（$C_N(\xi)$：$\xi$ により決る定数列）とする．このとき $N\to\infty$ で

$$(3\cdot18)\qquad \sum_{i=1}^{N}P_N(X_i\in E_N)=\xi+o(1).\quad (\xi>0,\ \text{const.})$$

$$(3\cdot19)\qquad \max_{D_{j+1}} P_N(X_{i_1}\in E_N \mid X_{i_2}\in E_N,\cdots,X_{i_{j+1}})=O\left(\frac{1}{N}\right),\ (j=1,\cdots,r),$$

を満すならば

$$(3.20) \qquad P_N(X_{N,k} \in E_N(\xi)) \sim \frac{1}{\Gamma(k)} \int_0^{\xi} t^{k-1} e^{-t}\, dt, \ (N \to \infty)$$

が成立する. ∎

謝辞 研究会の席上，有益な comments をくださった方々に深く感謝します.

REFERENCES

[1] 池田, 松縄 (1972). 極値統計量の一様漸近分布について. 数理解析研究所講究録 167.「統計的漸近理論」71-80.

[2] Ikeda, S. and T. Matsunawa, (1972). On the uniform asymptotic joint normality of sample quantiles. Ann. Inst. Statist. Math. 24 33-52.

[3] Chibisov, D. M. (1964). On limit distribution for order statistics. Theor. Probability Appl. 9 142-148.

[4] Meyer, R. M. (1967). Some Poisson-type limit theorems for sequences of dependent rare events, with applications. UNC Inst. Statist. Mimeo Series No. 529.

[5] Runnenburg, J. T. (1969). Limit theorems for stochastic processes occurring in studies of the light-sensitivity of the human eye. Statistica Neerlandica 23 1-17.

[6] Smirnov, N. V. (1967). Some remarks on limit laws of a variational series. Theor. Probability Appl. 12 337-339.

[7] Watson, G. S. (1954). Extreme values in sample from m-dependent stationary stochastic processes. Ann. Math. Statist. 25 798-800.