

# Stochastic ranking process の無限粒子極限と Amazon.co.jp のランキング

服部哲弥（東北大学・理），服部久美子（首都大学東京・数学）

Amazon.co.jp などのオンライン小売業のランキングの時間発展を記述する単純な多粒子系の確率モデルを提案する．このモデルの無限粒子極限（流体力学極限のように経験分布が分布値確率過程として決定論的な分布の時間発展に弱収束すること）は存在し，このモデルに基づく理論的予想は現実の web のデータと合うことを見つけた．結果として得られた経済学的なパラメータ（販売頻度をパレート分布で当てはめたときの指数）の値は，これまで分析されてきたよりもオンラインの小売業のインパクトが小さいことを意味している．いわゆる long tail 経済のありかたに整合する数理的解析手法として理論と実際両面の価値があると考えられる．

1. Stochastic ranking process . Stochastic ranking process  $\{X_i^{(N)}(t) \mid t \geq 0, i = 1, \dots, N\}$  は， $X_i^{(N)}(t)$  が粒子  $i$  の時刻  $t$  における順位（ランキング）を表す  $N$  粒子系の確率過程で，以下で定義する．各粒子  $i$  に対して自然数  $x_{i,0}^{(N)}$ （粒子  $i$  の初期順位  $X_i^{(N)}(0) = x_{i,0}^{(N)}$ ）と正数  $w_i^{(N)} > 0$ （jump 率）が与えられている．各  $i$  に対して増加確率変数列  $\tau_{i,j}^{(N)}, j = 1, 2, \dots$  は jump 時刻の列で，この時刻に粒子  $i$  は 1 位 ( $X_i^{(N)}(\tau_{i,j}) = 1$ ) になり，追い越された粒子  $i'$  の順位  $X_{i'}^{(N)}$  は順に繰り下がる： $X_{i'}^{(N)}(\tau_{i,j} - 0) < X_i^{(N)}(\tau_{i,j} - 0)$  なる  $i'$  に対して  $X_{i'}^{(N)}(\tau_{i,j}) = X_{i'}^{(N)}(\tau_{i,j} - 0) + 1$ ．Jump 時刻の列は粒子間で独立で各粒子ごとに独立増分で増分は指数分布  $P[\tau_{i,j+1}^{(N)} - \tau_{i,j}^{(N)} \leq t] = 1 - e^{-w_i^{(N)}t}$  に従うとする．

$x_C^{(N)}(t) = 1 + \#\{i \mid \tau_{i,1}^{(N)} \leq t\}$  の tail 側 ( $X_i^{(N)}(t) \geq x_C^{(N)}(t)$ ) の粒子は時刻  $t$  までに jump 未経験，head 側の粒子は jump 済み，であることに注意．主定理の極限の形も証明も両側で異なる．

$N$  が大きいときの振る舞いに興味がある．まず， $P[\tau_{i,1}^{(N)} \leq t] = 1 - e^{-w_i^{(N)}t}$  から大数の弱法則によって次を得る．

命題．Jump 率の分布  $\lambda^{(N)} = \frac{1}{N} \sum_{i=1}^N \delta_{w_i^{(N)}}$ （ $\delta_a$  は  $a$  に集中した単位分布）が  $N \rightarrow \infty$  で分布  $\lambda$  に

弱収束するならばスケールした軌道  $y_C^{(N)}(t) = \frac{1}{N} x_C^{(N)}(t)$  は  $y_C(t) = 1 - \int_0^\infty e^{-wt} \lambda(dw)$  に確率収束する． ◇

この軌道  $y_C(t)$  は実際の web のランキングの時間変化と良く合う（後述）．

2. 無限粒子スケール極限．直感的には，ランキングの上位には jump 率の大きい粒子が多く，下位は jump しにくい粒子が多いと考えられるが，流体力学極限のように経験分布の極限（従属確率変数の和の大数の法則）としてこの直感を厳密かつ精密に記述できる．以下を仮定する：

(1) 各  $y \in [0, 1)$  に対して jump 率の空間上の分布  $\mu_{y,0}$  が存在して， $N \rightarrow \infty$  で，スケールした初期ランキング  $y_{i,0}^{(N)} = \frac{1}{N} (x_{i,0}^{(N)} - 1)$  の分布  $\mu_{y,0}^{(N)}(dw dy) = \frac{1}{N} \sum_i \delta_{w_i^{(N)}}(dw) \otimes \delta_{y_{i,0}^{(N)}}(dy)$  は， $\mu_{y,0}(dw) dy$  に確率収束する．(2)  $\inf_{y \in [0,1)} \mu_{y,0}([0, M]) > 0$  なる  $M$  がとれる．(3)  $\lambda(\{0\}) = 0$ ．

仮定 (3) から直ちに  $y_C$  が狭義増加となるので逆関数  $t_0 : [0, 1) \rightarrow [0, \infty)$  が存在する． $y_C(t)$  の定義を拡張して  $y_C(y, t) = 1 - \int_y^1 \int_0^\infty e^{-wt} \mu_{z,0}(dw) dz$  とおく．仮定 (2) の下で  $y_C(y, t)$  は  $y$  について狭義増加となるので逆関数  $\hat{y}(\cdot, t) : [y_C(t), 1) \rightarrow [0, 1)$  が存在する．各  $(y, t) \in [0, 1) \times [0, \infty)$  に対して jump 率の空間上の分布  $\mu_{y,t}$  を以下で定義する：

$$\mu_{y,t}(dw) = \begin{cases} \frac{we^{-wt_0(y)} \lambda(dw)}{\int_0^\infty we^{-wt_0(y)} \lambda(dw)}, & y < y_C(t), \\ \frac{e^{-wt} \mu_{\hat{y}(y,t),0}(dw)}{\int_0^\infty e^{-wt} \mu_{\hat{y}(y,t),0}(dw)}, & y > y_C(t). \end{cases}$$

主定理 . (1)(2)(3) を仮定すると , スケールしたランキング  $Y_i^{(N)}(t) = \frac{1}{N} (X_i^{(N)}(t) - 1)$  の位置と jump 率の直積空間上の経験分布 (分布値確率変数)

$$\mu_{y,t}^{(N)}(dw dy) = \frac{1}{N} \sum_i \delta_{w_i^{(N)}}(dw) \otimes \delta_{Y_i^{(N)}(t)}(dy)$$

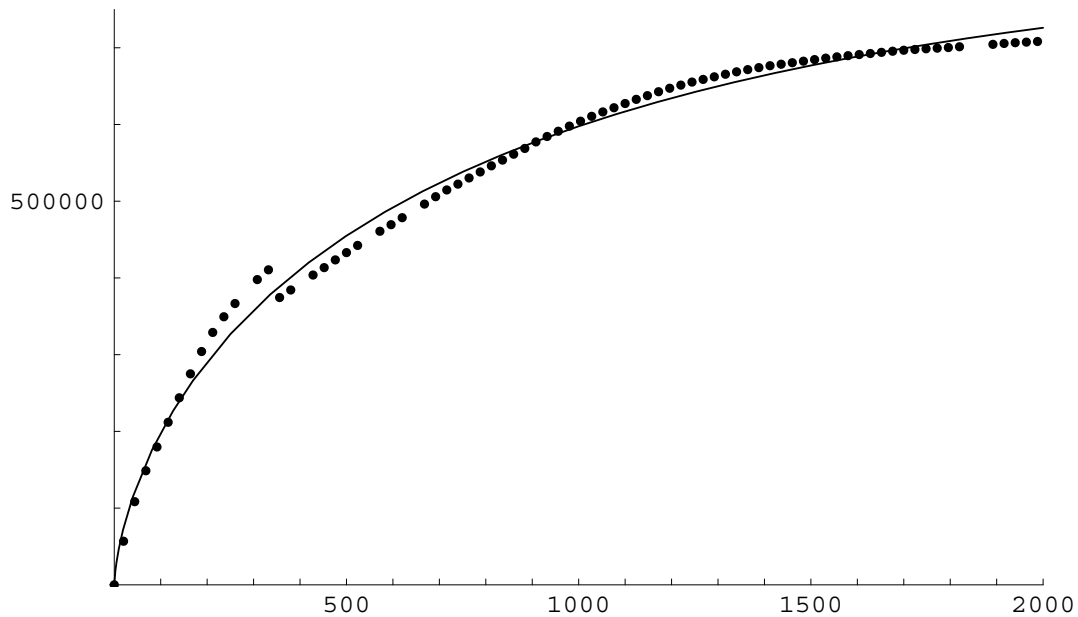
は  $N \rightarrow \infty$  で  $\mu_{y,t}(dw) dy$  に確率収束する .

◇

$\mu_{y,t}(dw)$  は次の非線形偏微分方程式系の解である :

$$\frac{\partial \mu_{y,t}(dw)}{\partial t} + \frac{\partial (v(y,t) \mu_{y,t}(dw))}{\partial y} = -w \mu_{y,t}(dw), \quad \frac{\partial v}{\partial y}(y,t) = - \int w \mu_{y,t}(dw).$$

3 . Online ranking 分析への応用 . Stochastic ranking process は Amazon.co.jp の本のランキングの時間変化の数学モデルと見ることができる . モデルとするには jump 率の分布  $\lambda$  を決める必要があるが , 社会学や経済学では Pareto 分布 (log-linear 分布) が用いられることが多いようである . たとえば  $i$  番目に高所得の人の所得は  $w_i = a (N/i)^{1/b}$  で与えられるとする分布のモデルである . これを用いると  $t = 0$  で 1 位の粒子 (本) の時刻  $t$  の順位は  $x_C(t) \simeq N (1 - b(at)^b \Gamma(-b, at)) + 1$  となる .  $\Gamma(z, p) = \int_p^\infty e^{-w} w^{z-1} dw$  は不完全ガンマ関数 .



図の点は Amazon.co.jp のランキングの実測データ , 曲線は  $x_C(t)$  を当てはめたものである (横軸は時間 , 縦軸は順位) . パラメータ  $N, a, b$  と初期時刻の 4 個を 77 個のデータに対して最小 2 乗法で当てはめた . 特に指数  $b \simeq 0.6$  を得た . 既存の online retail 分析では古典的な観測手法で得られた  $b > 1$  が用いられているが , これは 'long-tail' 側を過大評価している可能性がある . 我々の方法において経済学的に重要な指数  $b$  をデータから得る背景には , ランキング上の 1 粒子の軌道によって tail 側の多数の粒子 (売れ行きの小さい多数の商品) の総和の動向が正確に記述できるという数学的裏付け (大数の法則) がある .

参考文献 .

[1] K. Hattori, T. Hattori, *Existence of an infinite particle limit of stochastic ranking processes*, preprint (2007).

[2] K. Hattori, T. Hattori, *Equation of motion for incompressible mixed fluid driven by evaporation and its application to online rankings*, preprint (2007).

いずれも <http://www.math.tohoku.ac.jp/~hattori/liamazn.htm> に置いてあります .