Stochastic ranking process の流体力学極限と 2ch.net

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り業のランキングの時間発展を記述する多粒子系の確率モデルを提案し，その無限粒子極限が存在すること，つまり，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，を証明する．この定理に基づく理論的予想が現実の web のデータと良く合うことを実証し，結果として得られる社会学的なパラメータ（パレート分布の指数）の実測値が，これまで計量経済学的に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する，ということにも言及する．

１. Stochastic ranking process． Stochastic ranking process $\{X_i^{(N)}(t) \mid t \geqq 0,\ i=1,\cdots,N\}$ は，$X_i^{(N)}(t)$ が粒子 i の時刻 t における順位を表す N 粒子系の確率過程で，以下で定義する．各 i に対して自然数 $x_{i,0}^{(N)}$（粒子 i の初期順位 $x_{i,0}^{(N)} = X_i^{(N)}(0)$）と正数 $w_i^{(N)} > 0$ (jump rate) が与えられている．各 i に対して増加確率変数列 $\tau_{i,j}^{(N)}$, $j=0,1,2,\cdots$, は jump time の列で，この時刻に粒子 i は 1 位 $X_i^{(N)}(\tau_{i,j}) = 1$ になり，他の粒子の順位 $X_{i'}^{(N)}$, $i' \neq i$ は順に繰り下がって，全体として 1 位から N 位の順位を保つ．Jump time の列は粒子間で独立で各粒子ごとにも独立増分でその分布は指数分布 $\mathrm{P}[\,\tau_{i,j+1}^{(N)} - \tau_{i,j}^{(N)} \leqq t\,] = 1 - e^{-w_i^{(N)}t}$ とする．

$x_C^{(N)}(t) = \sharp\{i \mid \tau_i^{(N)} \leqq t\}$ の tail 側 $(X_i^{(N)}(t) > x_C^{(N)}(t))$ の粒子は時刻 t までに jump 未経験，head 側の粒子は jump 済み，となることに注意．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，$\mathrm{P}[\,\tau_i^{(N)} \leqq t\,] = 1 - e^{-w_i^{(N)}t}$ から大数の弱法則によって次を得る．

命題． Jump rate の分布 $\lambda^{(N)}(dw) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \delta(w - w_i^{(N)})\,dw$ が $N \to \infty$ で分布 λ に弱収束するならばスケールされた軌道 $y_C^{(N)}(t) = \dfrac{1}{N} x_C^{(N)}(t)$ は $y_C(t) = 1 - \displaystyle\int_0^\infty e^{-wt}\lambda(dw)$ に確率収束する．$\diamond$

この軌道 $y_C(t)$ は実際の web のランキングと良く合う（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 rate の大きい粒子が多く，下位は jump しにくい粒子がたまりやすいと考えられるが，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) $N \to \infty$ で，スケールされた初期ランキング $y_{i,0}^{(N)} = \dfrac{1}{N}\,(x_{i,0}^{(N)} - 1)$ の分布（分布値確率変数）

$$\mu_{y,0}^{(N)}(dw)\,dy = \frac{1}{N}\sum_i \delta(w - w_i^{(N)})\delta(y - y_{i,0}^{(N)})\,dw\,dy$$

は y について一様に，y のルベーグ測度に関して絶対連続な分布 $\mu_{y,0}(dw)\,dy$ に確率収束する．

(2) $\displaystyle\inf_{y\in[0,1)} \mu_{y,0}([0,M]) > 0$ 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 $t_0:\ [0,1) \to [0,\infty)$ が存在する．$y_C(t)$ の定義を拡張して $y_C(y,t) = 1 - \displaystyle\int_y^1 \int_0^\infty e^{-wt}\mu_{z,0}(dw)\,dz$ 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 $\hat{y}(\cdot,t):\ [y_C(t),1) \to [0,1)$ が存在する．

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

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

は $N \to \infty$ で，y 空間上のルベーグ測度に関する密度

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

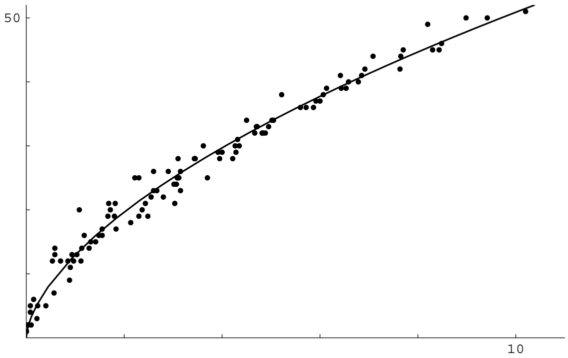
を持つ分布 $\mu_{y,t}(dw)\,dy$ に確率収束する． ◇

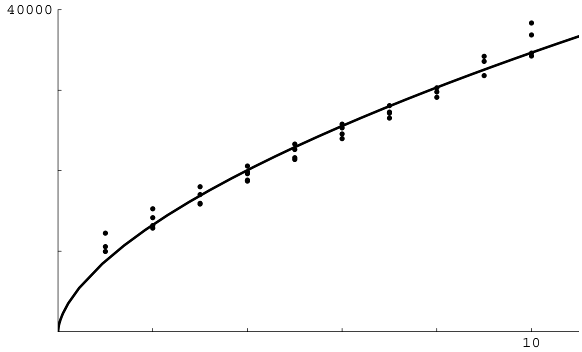
主定理は，従属確率変数の和の大数の法則という意味で自明ではない．

$\mu_{y,t}(dw)$ は次の偏微分方程式系の解である：

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

３．2ch.net と Amazon.co.jp． Stochastic ranking process は 2ch.net のスレッド一覧の時間変化や 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 rate の分布 λ 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 $w_i = a\left(\dfrac{N}{i}\right)^{1/b}$ 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（本やスレッド）の時刻 t の順位は $x_C(t) = N\,(1 - b(at)^b\Gamma(-b, at)) + 1$ となる．$\Gamma(z,p) = \displaystyle\int_p^\infty e^{-w}w^{z-1}\,dw$ は不完全ガンマ関数．左図と右図の点はそれぞれ 2ch.net と 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．（いずれも横軸は時間，縦軸は順位.）パラメータ a, b の 2 個（Amazon では N も）を fit するだけで驚くほどよく合う．





特に指数 b の最尤値はいずれも $b \simeq 0.6$ となった．既存の online retail 分析では古典的な手法で得られた $b > 1$ が用いられているが，これは ‘long-tail’ 側を過大評価することになり，online retail の影響を過大評価していると考えられる．

我々の方法は応用上重要な指数 b をデータから直接得る．その背景には主定理のように，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（応用上は売れ行きの小さい多数の商品）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．

また，2ch.net は stochastic ranking process の定義に近い単純な構造を持っており，スレッド（粒子）の jump の時刻がレスとして全て記録・公開されている．E. Coli が単純な生命として分子生物学の初期の基礎技術確立に大きな役割を果たしたと聞くが，同様に 2ch.net は long tail の経済学の基礎技術に役立つと考えられる．

参考文献．

[1] K. Hattori, T. Hattori, *Existence of 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．