

Amazon ランキングと確率順位付けモデルの流体力学 極限

服部哲弥 (慶應義塾大学)*

1. 積み上げた本の恨み

ウェブで見られるランキング (流行度の順位付け) という 21 世紀的な社会現象を, その時間変化の理論的考察という切り口で紹介する.

現代の現象を後回しにして, 理論の歴史に先に言及する. 遡ること約半世紀 Tsetlin が (定常的なノイズの下の有限状態オートマトンの例として) 机の上に積み上げた本の塔から毎回必要な本を抜き出して参照を終えたら塔の一番上に置くという規則を記している [18]. 定常状態のあらわな公式を導いて (期待されるとおり) たびたび参照する本ほど先頭近く (塔の上部) にいる確率が高いことを指摘した上で, その節の最後を「机上の本が (せっかく使いやすい定常分布に落ち着いていたのに, 片付けられて) 順番に並び直されたときのがっかりを説明するかもしれない」という文で締めていて, かつて自分の勉強机という聖域を侵された鬱憤を晴らすために選んだ例に見える.

本の電子化が進んだ半世紀後の現在, Tsetlin の積年の恨みと積ん読の擁護にぴんと来ない世代も増えているかもしれないが, 先頭に跳ぶ規則 (move-to-front rule) と名付けられたこの規則は 20 世紀終盤に計算機のデータの動的配置 (計算実行時の使用状況に応じたデータの記憶装置内の配置) のアルゴリズムの 1 つとして研究が流行した. 磁気テープのように記憶媒体先頭からの相対的な位置に基づいて (つまり, 先頭に巻き戻して) データを探す記憶装置では, 頻繁に使われるデータを先頭近くに置きたいが, データを使ったらそのまま先頭に配置し, 直前まで先頭以下の好順位を占めていたデータたちを順に下位に移動すれば, Tsetlin の主張する意味で自然な記憶装置内のデータ配置を実現する.

以下, 本やデータを抽象化して粒子と呼び, 多粒子系の問題として扱う. 先頭に跳ぶ規則では, 先頭に跳ばなかった粒子についてもその順序情報を更新する必要がある. 机上の本では下のほうの本を引っっこ抜けばその上の本は重力で自動的に落ちるが, 記憶装置のデータは順序情報を書き直さねばならないので貴重な計算時間を消費する. 半導体記憶装置の時代になって, 値段と速さが相反する記憶媒体系にデータを振り分けるキャッシュのモデル (least-recently-used (LRU) caching) に引き継がれたが, 高速化のぎりぎりのところを競う計算機の動作に対して使わないデータの順位情報書き直しの時間的コストを考えると, LRU キャッシングは実用というよりは粒子数 N を大きくした極限があらわに書けることに依拠した理論的興味が勝ってきたかもしれない.

ここで先頭に跳ぶ規則を現代的に整理した数式で書く. $T > 0$ を以後固定する. 自然数 N (粒子数) を与えるごとに本稿で扱う方程式に現れる $i = 1, \dots, N$ と $t \in [0, T]$ に対する (標準の確率空間 (Ω, \mathcal{F}, P) 上の) 確率変数 $Y_i^{(N)}(t) : \Omega \rightarrow [0, 1]$ を, 図 1 のキャ

本研究は現在科研費基盤研究 (C)18K03344 の支援を受けています.

2010 Mathematics Subject Classification: 60K35, 82C22

キーワード: 確率的順位, 直前の到着時刻に強度が依存する点過程, 流体力学極限, 大数の完全法則

* 〒 223-8521 横浜市港北区日吉 4-1-1 慶應義塾大学経済学部日吉数学研究室

e-mail: hattori@econ.keio.ac.jp

web: <http://web.econ.keio.ac.jp/staff/hattori/hattori.htm>

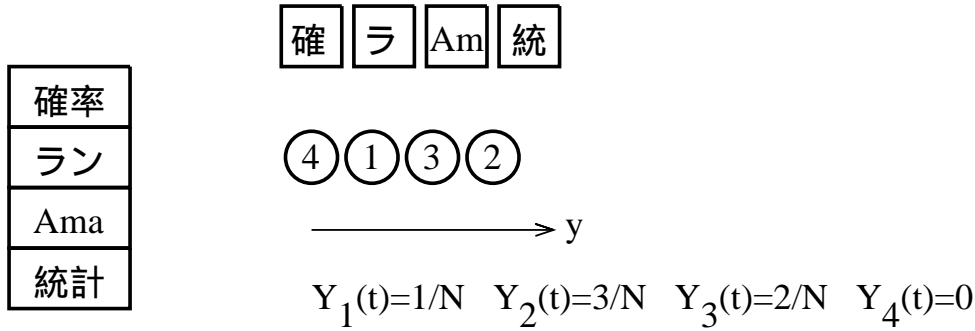


図 1: 図左: 机上に積み上げた本. 図右上: 行数節約のため横倒し (左が先頭, 右が下位). 図右中: 題名は番号に, 座標を入れて順位は位置と呼ぶ. 図右下: 番号 (題名) を添字, 位置 (順位) を値とする確率変数で定式化 (N は粒子数).

プシヨンの手順で, 時刻 t での粒子 i の位置と扱う. 各粒子 i について $\tilde{v}_i^{(N)}$ を非負実数値 w_i を強度とする i について独立なポワソン過程として, 初期順位 $Y_i^{(N)}(0) = y_i^{(N)}$ も固定し, 永幡幸生さん [16] に従って

$$Y_i^{(N)}(t) = y_i^{(N)} + \frac{1}{N} \sum_{j=1}^N \int_0^t \mathbf{1}_{Y_j^{(N)}(s-) > Y_i^{(N)}(s-)} \tilde{v}_j^{(N)}(ds) - \int_0^t Y_i^{(N)}(s-) \tilde{v}_i^{(N)}(ds) \quad (1)$$

で一意に決まる粒子系の確率的な時間発展を考える. ここで, 集合 A 上で 1 で補集合 A^c 上で 0 なる関数を $\mathbf{1}_A$ と書き, また, 時刻 s の関数 $Y_i^{(N)}(s)$ において s についての左極限の値を $Y_i^{(N)}(s-)$ と書く. また, 強度 w のポワソン過程 \tilde{v} はサンプル毎に非減少で増分 ($t > s$ のときの $\tilde{v}(t) - \tilde{v}(s)$) が平均 $w \times (t - s)$ のポワソン分布

$$P[\tilde{v}(t) - \tilde{v}(s) = k] = e^{-w(t-s)} \frac{w^k (t-s)^k}{k!}, \quad k \in \mathbb{Z}_+, \quad (2)$$

に従い, 重ならない時間の増分が独立な確率変数であるものを言う. (1) で定まる粒子系を確率的順位付け模型と呼ぶ.

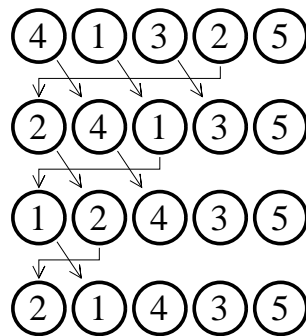


図 2: ポワソン過程 $\tilde{v}_i^{(N)}$ の k 回目の到着時刻 $\tau_{i,k}$ が $\tau_{2,1} < \tau_{1,1} < \tau_{2,2} < \dots$ の例.

(1) の右辺第 3 項を説明するために, 時刻 τ が $\tilde{v}_i^{(N)}$ の到着時刻, すなわち $\tilde{v}_i^{(N)}(\tau) - \tilde{v}_i^{(N)}(\tau-) = 1$, とすると (1) の $t = \tau$ と $t = \tau-$ の差は

$$Y_i^{(N)}(\tau) - Y_i^{(N)}(\tau-) = -Y_i^{(N)}(\tau-) \int_{\tau-}^{\tau} \tilde{v}_i^{(N)}(ds) = -Y_i^{(N)}(\tau-)$$

なので $Y_i^{(N)}(\tau) = 0$, すなわち $\tilde{v}_i^{(N)}$ の到着時刻 τ は粒子 i が先頭に跳ぶ時刻である . したがってさらに , 右辺第 2 項は注目する粒子 i より下位の j が先頭に跳んだ影響で順位を下げることを意味する (図 2) . この意味で (1) は先頭に跳ぶ規則の定式化である .

2. Amazon はロングテールに非ず

時は 21 世紀 . 本が参照の多い順に並ぶ Tsetlin の考察が電子化され , 人気もしくは流行の度合いの大規模な順位がリアルタイムで更新されるのを目の当たりにする時代になった .



図 3: Amazon.co.jp の本のページ (紙面の都合で実際よりも大幅に省略した . ランキングは , 2019 年 4 月現在 「登録情報」 の項の中 (図で最下部の囲んだところ) にあり , 毎時 1 回更新される .

計算機とネットワークの爆発的發展によって成立した今世紀特有の現象であるネット小売業の成功例 Amazon.com は , ネット書店から出発した . その日本法人の本のページには図 3 のように 「ランキング」 と名付けられた時間変化する順位が表示されている . Amazon.co.jp は最新販売数 , 言い換えると , 各時点の流行度あるいは人気度を考慮して順位づけているとしつつ , この順位の正確なアルゴリズムを秘匿している . 秘匿された Amazon のアルゴリズムを見抜くことは目指さず , 逆転の発想で , 流行度のもっとも単純な定義を考えた結果 , (1) の $\tilde{v}_i^{(N)}$ を本 i の注文数と解釈して確率順位付け模型 [4, 5, 6, 8, 7] に至った . そしてそれが実際の Amazon によるランキングをどの程度説明するか実測することを考えた .

$Y_i^{(N)}$ たちは (1) を見ると従属確率変数列なのでそのままでは計算できないが , 先頭に跳ぶ第 3 項を落として初期値 $\gamma = (y_0, t_0) \in [0, 1] \times [0, T]$ と $t \geq t_0$ の関数とした

$$Y_C^{(N)}(\gamma, t) = y_0 + \frac{1}{N} \sum_{j=1}^N \int_{t_0}^t \mathbf{1}_{Y_j^{(N)}(s-) > Y_C^{(N)}(\gamma, s-)} \tilde{v}_j^{(N)}(ds) \quad (3)$$

は右辺の形から時刻 t_0 に y_0 より下位にあった粒子に付随したポワソン過程のうち時間 $(t_0, t]$ 中に到着があった個数を数えるだけなので，独立確率変数の平均だから大数の強法則が成り立つ．簡単のため， $\gamma = (y_0, t_0) = (0, t_0)$ の場合を考えて

$$y_C(\gamma, t) = 1 - \int_W e^{-w(t-t_0)} \lambda(dw), \quad t \geq t_0, \quad (4)$$

と置く． δ_c を 1 点 c に集中した単位分布として，強度 $w_i \in W := \mathbb{R}_+$ の分布を

$$\lambda^{(N)}(dw) = \frac{1}{N} \sum_{i=1}^N \delta_{w_i} \quad (5)$$

と置く． $y_0 = 0$ のとき以上の考察と (2) と (5) から

$$\begin{aligned} E[Y_C^{(N)}((0, t_0), t)] &= \frac{1}{N} \sum_{j=1}^N P[\tilde{v}_j(t) - \tilde{v}_j(t_0) > 0] = \frac{1}{N} \sum_{j=1}^N (1 - e^{-w_j(t-t_0)}) \\ &= 1 - \int_W e^{-w(t-t_0)} \lambda^{(N)}(dw) \end{aligned}$$

であることと，大数の強法則（細かく言えば大数の完全法則 [13, 11]）から，

$$\lim_{N \rightarrow \infty} Y_C^{(N)}((0, t_0), t) = y_C((0, t_0), t), \quad a.e., \quad (6)$$

を得る [4]． $Y_C^{(N)}$ の定義 (3) は，粒子の位置変化 $Y_i^{(N)}$ の定義 (1) の「先頭に跳ぶ」第 3 項を落としたものなので，並んでいる粒子の間に挟んだ先頭に跳ばない「しおり」の動きである．言い換えると，粒子は先頭に跳ばない間は（初期値をそろえたとき） $Y_C^{(N)}$ と同じ時間発展をする．

(4) が実際の Amazon ランキングをどの程度説明するか実測と突き合わせるために，まず，図 1 で順位を位置に直す際に $Y_i^{(N)}$ は値が $[0, 1)$ に収まるように規格化したことを思い出すと順位は $X_i^{(N)}(t) = NY_i^{(N)}(t) + 1$ となる．Amazon.co.jp の本は研究開始当時に N が百万 (10^6) の程度，現時点でその約 2 倍と推測されるので $N \rightarrow \infty$ の極限 (6) は良い近似として (4) を用いる．(1) の $\tilde{v}_i^{(N)}$ を本 i の注文数と解釈する独自の着想に基づいて，本 i の累積 k 冊目の注文時刻を $\tau_{i,k}$ として， $\tau_{i,k} \leq t < \tau_{i,k+1}$ において

$$X_i^{(N)}(t) = N \int_0^\infty (1 - e^{-w(t-\tau_{i,k})}) \lambda(dw) + o(N) \quad (7)$$

を得る．時刻 t の直近で本 i が売れた時刻 $\tau_{i,k}$ を時刻の起点に取り直せば（ $\tau_{i,k}$ で条件付ければ）右辺がバラツキのない量なのは，多数の刊行書の売り上げのバラツキの相殺という，大数の法則であり，多数の本の売上の総計で相対的に順位を落とすので，どの本に注目しても（ i によらず）同形の曲線に従う．Amazon ランキングと理論式 (7) を比べるにはもう 1 つ， $W = \mathbb{R}_+$ 上の分布 λ が必要である．これは強度分布すなわち各書籍の平均売上の分布と翻訳されるが（ここでも企業の種々の詳しいデータは秘匿されているので）分布の概形（分布族）を指定して，パラメータ（母数）をデータから統計的に当てはめる（年齢などの属性を適切に揃えたときの）身長分布のように平均とそのまわりの揺らぎの描像で語られる場合は正規分布族，所得などのように格差が議論となるときはべき分布（一般化パレート分布，離散版 λ_N では一般化ジップの法

則) 族が広く使われる．専門書と流行書の売り上げ格差を思い浮かべるまでもなくここでは後者を選んで，

$$\lambda([w, \infty)) = \left(\frac{a}{\max\{w, a\}} \right)^b \quad (8)$$

と置く． a と b は正定数である．(8)を(7)に代入すると，

$$X_i^{(N)}(t) \sim N - Ne^{-a(t-\tau_{i,k})} + N(a(t-\tau_{i,k}))^b \Gamma(1-b, a(t-\tau_{i,k})) \quad (9)$$

と，不完全ガンマ関数 $\Gamma(z, p) = \int_p^\infty e^{-x} x^{z-1} dx$ を用いて理論曲線が求まる．図4のデー

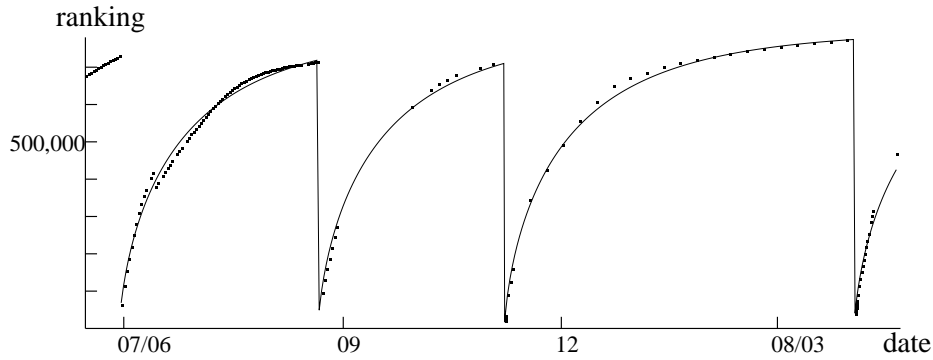


図4: 横軸は時刻 t ，全体で約1年，縦軸は X_i ，図の下方横軸近くが小さな値すなわち上位．点はランキングの時間変化の観測値．左のほうの密な点列は毎日定時の観測，右のほうは週1度の観測．曲線は確率的順位付け模型の理論曲線の観測値への当てはめ．

タは筆者が初めて実測した Amazon ランキングの時間発展の記録である [4, 5, 6, 8, 7]．順位が改善するときはそれまで落としていた順位が一気に1位付近(横軸近く)まで跳ぶ．図4のデータから $\tau_{i,k}$ たちとパラメータ N, a, b を求めることで得た

$$(N, a, b) \sim (8 \times 10^5, 5 \times 10^{-4}, 0.8) \quad (10)$$

を(9)に代入したのが図4の曲線である．

「注文を受けた瞬間は人気度または流行度1位」という，これ以上なく単純な理論が図4の点たちと曲線の関係程度には現実のランキングのデータを説明していることに気をよくして，Tsetlinの定常分布(の，粒子数 N を大きくした極限)についても考察する．粒子系の位置強度結合経験分布を

$$\mu_t^{(N)} = \frac{1}{N} \sum_{i=1}^N \delta_{(w_i, Y_i^{(N)}(t))} \quad (11)$$

と置き， $t=0$ での分布 $\mu_0^{(N)}$ の N についての列を， $N \rightarrow \infty$ とともに弱収束するように選んで，極限を $\mu_0 = \lim_{N \rightarrow \infty} \mu_0^{(N)}$ とおく(特に(5)から $\lambda^{(N)}(dw) = \mu_t^{(N)}(dw \times [0, 1])$ は時刻によらず $N \rightarrow \infty$ で $\lambda(dw) = \mu_0(dw \times [0, 1])$ に収束する.) $Y_i^{(N)}$ と同様に $\mu_t^{(N)}(dw \times [y, 1])$ も従属確率変数の和だが， $\mu_t^{(N)}$ の位置についての分布関数である測度値確率過程

$$\varphi^{(N)}(dw, \gamma, t) = \mu_t^{(N)}(dw \times [Y_C^{(N)}(\gamma, t), 1]), \quad \gamma = (y_0, t_0) \in [0, 1] \times [0, T], \quad t \geq t_0, \quad (12)$$

は上手な組み合わせで，時刻 t_0 において y_0 より下位にいた粒子 i たちのうちで独立なポワソン過程 $\tilde{v}_i^{(N)}$ の到着時刻が時間 $(t_0, t]$ 中にあるものの割合だから独立確率変数列の大数の完全法則で極限の存在がわかり， $\mu_t^{(N)}$ もある分布 μ_t に収束する．特に $y_0 = 0$ のとき (4) と同様に，

$$\mu_t(dw \times [y_C(\gamma, t), 1]) = \lim_{N \rightarrow \infty} \varphi^{(N)}(dw, \gamma, t) = e^{-w(t-t_0)} \lambda(dw), \quad \gamma = (0, t_0), \quad t \geq t_0,$$

を得る．ここで増加関数 $y = y_C((0, 0), t)$ の逆関数を $t_C : [0, 1) \rightarrow \mathbb{R}_+$ と置くと，(4) から $t_C(y_C(\gamma, t)) = t - t_0$ および $y = 1 - \int_W e^{-wt_C(y)} \lambda(dw)$ であって，極限の定常分布（時刻によらない分布）は

$$\mu_t(dw \times [y, 1]) = \mu_0(dw \times [y, 1]) = e^{-wt_C(y)} \lambda(dw), \quad (13)$$

あるいは，微分 $t'_C(y_C) y'_C = 1$ に (4) を使った後に $y_C = y$ と $t - t_0 = t_C(y)$ で書き直して

$$\mu_0(dw \times dy) = \frac{we^{-wt_C(y)}}{\int_0^\infty ve^{-vt_C(y)} \lambda(dv)} \lambda(dw) dy$$

であることもわかる [4, 6, 7] .

Amazon ランキングの実測と先頭に跳ぶ規則あるいは確率順位付け模型の数学のその他の関係について簡単に列挙する．まず，Amazon ランキングは同社が恣意的に定めたアルゴリズムに基づく数値であって，先頭に跳ぶ規則を遵守していると信じる理由はなく，ここに確率順位付け模型を当てはめたのは筆者による．その後の継続的なデータ収集によって先頭に跳ぶ規則からの逸脱が見つかっている．それでも図 4 の縮尺ではデータ（点）と理論（曲線）は合う．その理由を考えると，まれにしか売れない本たちの間で無理に流行度による順位を付けるとすると（直前の売り上げより以前の売り上げは，現在の流行と無関係なくらい遠い昔なので）直前の売り上げが新しいほど現在の流行度は高いとするしかないので，流行度に基づく順位をどう定義しても売れていない本の間の順位は最後に売れた順になる [7] . 最後に売れた順位は先頭に跳ぶ規則に基づく順位と等しい．他方，世の中の大半の本は殆ど売れていないから，Amazon ランキングの（上位の一部を除く）大半の順位は，Amazon が流行度をきちんと考慮してアルゴリズムを決める限り先頭に跳ぶ規則が悪くない．

次に，社会現象や自然現象の活動の昼夜差については，ポワソン過程 $\tilde{v}_i^{(N)}$ の強度 w_i を時刻の関数に一般化して非一様ポワソン過程とすれば，独立増分性が成り立つので，確率論として変わることなく確率順位付け模型の極限の確率を（非一様）ポワソン過程の確率で書ける．たとえば，軌道 (4) で積分範囲 W を \mathbb{R}_+ から関数の集合に置き換え， $w(t - t_0) \mapsto \int_{t_0}^t w(s) ds$ と置き換えた式が成り立つ．強度が時刻依存性を持つ現実の Amazon ランキングのデータを定数強度に基づく (9) に当てはめて良いことの理屈も難しくなく [3] . 経験分布の収束だけでなく名札付き粒子過程の独立確率過程列への収束（カオスの伝搬）が成り立ち [16] ，さらに，汎関数中心極限定理の成立も永幡さんによってわかっている [17] .

最後に，商品の売上分布は社外秘に属するが，ランキングという公開情報だけを用いることで Amazon.co.jp の売上分布の近似である (8) と (10) を得た．例えばロングテール

ビジネスであるか否かを部外者でありながら分析できる．ネット書店としての Amazon は、かつてロングテールビジネスの草分け的存在として注目された [1]．ビッグヒットを除く裾野に位置する本は個別にはめったに売れなくても、ほとんどすべての本が裾野に位置するので、合計すれば無視できない売上をもたらすかもしれない、というのがロングテールビジネスの可能性である．分布の形 (8) のとおり、 b が小さいとき裾野に比べてビッグヒットの寄与が圧倒的であり、 b が大きいときはロングテールが無視できない． N が大きいとき、ロングテールの売上への寄与が全売上の中で無視できるかできないかは b が 1 より大きいか小さいかが判定基準となることがわかる．データを当てはめた結果 (10) から $b < 1$ とわかったので、Amazon.co.jp ではロングテールの売上は無視できる．Amazon はロングテールに非ずである [7] ．

3. 強度が位置依存性を持つ確率的順位付けモデルの流体力学極限

初期の研究の頃にセミナー等で必ず質問されたことの 1 つに、ランキングが上位であることの宣伝効果はあるか、ということがあった．この問題に数学的に答えるには強度 $w(x, t)$ が時刻 t だけでなく位置 x にも依存する場合を考える必要がある．

第 1 象限 $\mathbb{R}_+ \times \mathbb{R}_+$ 上の単位強度のポワソン乱測度 $\nu_i^{(N)}$ の独立な列 $i = 1, \dots, N$ を取る．すなわち、 $A \subset \mathbb{R}_+^2$ に対して $U = \nu_i^{(N)}(A)$ は平均が面積 $(E[U([a, b] \times [c, d])]) = (b - a)(d - c)$ のポワソン分布に従う確率変数で、排反な集合 $A, B \subset \mathbb{R}_+^2$ に対して $U(A)$ と $U(B)$ が独立とする．楠岡誠一郎さん [15] に従って、(1) において各 i に対して $\tilde{\nu}_i^{(N)}$ を

$$\tilde{\nu}_i^{(N)}(t) = \int_{s \in (0, t]} \int_{\xi \in \mathbb{R}_+} \mathbf{1}_{\xi \in [0, w_i(Y_i^{(N)}(s-), s))} \nu_i^{(N)}(d\xi \times ds) \quad (14)$$

に置き換えた確率微分方程式系を強度が位置依存性を持つ確率順位付けモデルと呼ぶ． $w_i(y, s)$ が定数ならば (14) の $\tilde{\nu}_i^{(N)}$ は、 $\tilde{\nu}_i^{(N)}(t) = \nu_i^{(N)}([0, w_i] \times (0, t])$ なので、 $\tilde{\nu}_i^{(N)}$ は強度 w_i のポワソン過程であり、第 2 節で紹介した確率順位付けモデルの原型に戻るが、一般には (14) はその形から独立増分でもなく i についての独立性もないので難しくなる．

位置強度結合経験分布 (11) はばらつく (サンプル $\omega \in \Omega$ の関数である) が、初期値 $\mu_0^{(N)}$ が $N \rightarrow \infty$ でサンプルによらない分布 μ_0 に収束すれば、サンプルによらない分布 μ_t の時間発展に概収束する、ということ (流体力学極限) が主定理である．主定理の仮定の詳細は原著および最近書く機会を頂いた解説 [14] に譲るが、概略、強度関数の分布 $\lambda^{(N)}$ については強度のノルムが収束し、強度関数の空間微分は有界で、初期時刻 $t = 0$ での位置強度結合分布 $\mu_0^{(N)}$ は収束する、という自然な仮定の下で、以下が成り立つ．

定理 1 (主定理, [12, Theorem 2]) $\mu_0^{(N)}$ の極限を μ_0 と置くととき、確率 1 で $N \rightarrow \infty$ のとき (11) の $\mu_t^{(N)}$ は t について一様にサンプルによらない μ_t に弱収束する．すなわち、任意の $y \in [0, 1]$ と有界連続関数 $h: W \rightarrow \mathbb{R}$ に対して

$$\lim_{N \rightarrow \infty} \sup_{t \in [0, T]} \left| \int_W h(w) \mu_t^{(N)}(dw \times [y, 1]) - \int_W h(w) \mu_t(dw \times [y, 1]) \right| = 0, \text{ a.e.,}$$

が成り立つ．

さらに、カオスの伝搬 (名札付き部分有限粒子系の独立確率過程への収束, propagation of chaos) が成り立つ．すなわち、正整数 L と $y_i \in [0, 1)$, $i = 1, 2, \dots, L$, に対

して

$$i = 1, 2, \dots, L, \text{ について } \nu_i^{(N)} = \nu_i, \quad N \in \mathbb{N}, \text{ および } \lim_{N \rightarrow \infty} y_i^{(N)} = y_i,$$

が成り立つとき, 名札付き粒子系 $(Y_1^{(N)}(t), Y_2^{(N)}(t), \dots, Y_L^{(N)}(t))$ は確率 1 で $N \rightarrow \infty$ のとき $t \in [0, T]$ について一様に $(Y_1(t), Y_2(t), \dots, Y_L(t))$ に収束する. ここで, 各 $i = 1, 2, \dots, L$ に対して Y_i は

$$\begin{aligned} Y_i(t) &= y_i + \int_{s \in (0, t]} \int_{(w, z) \in W \times [Y_i(s-), 1]} w(z, s) \mu_s(dw \times dz) ds \\ &\quad - \int_{s \in (0, t]} \int_{\xi \in \mathbb{R}_+} Y_i(s-) \mathbf{1}_{\xi \in [0, w_i(Y_i(s-), s))} \nu_i(d\xi \times ds), \\ & \quad i = 1, 2, \dots, L, \quad t \in [0, T], \end{aligned}$$

の唯一解である. ◇

点過程 (非減少右連続非負整数値初期値 0 の確率過程) $\tilde{\nu}$ であって, $\tau_0 = 0$ として, 各正整数点 k への到着時刻 $\tau_k = \inf\{t \geq 0 \mid \tilde{\nu}(t) \geq k\}$ の, 直前の到着時刻 τ_{k-1} についての条件付き確率が $t \geq \tau_{k-1}$ に対して $P[t < \tau_k \mid \mathcal{F}_{\tau_{k-1}}] = \exp(-\int_{\tau_{k-1}}^t w(\tau_{k-1}, u) du)$ で与えられるものを, $w(s, t)$ を強度とする直前の到着時刻に強度が依存する点過程と名付ける [9, 10]. 応用上は, 照明が性能に応じた強度関数に従ってランダムに切れるが, 切れた時点で買い換えると, 技術の進捗や環境上の規制によって, その時点で買える照明の規格が違ふことを考慮した累積交換回数を表す. ポワソン過程とは対照的に, 独立増分ではなく, 公式は複雑になる. たとえば, 時間 $(s, t]$ に到着が無い確率は, 一時的に $W(u, v) = \int_u^v w(u, s) ds$ とおくと, $k = 1, 2, 3, \dots$ について, $u_0 = 0$ として

$$P[\tilde{\nu}(t) = \tilde{\nu}(s) = k] = \int_{0 \leq u_1 \leq u_2 \leq \dots \leq u_k \leq s} e^{-W(u_k, t)} \prod_{i=1}^k w(u_{i-1}, u_i) e^{-W(u_{i-1}, u_i)} du_i$$

となる [9, §3]. ここで, w が第 1 変数について定数ならば, よく知られた公式

$$\int_{0 \leq u_1 \leq u_2 \leq \dots \leq u_k \leq s} \prod_{i=1}^k f(u_i) du_1 du_2 \dots du_k = \frac{1}{k!} \left(\int_0^s f(v) dv \right)^k, \quad s \geq 0, \quad k = 1, 2, \dots,$$

によって $P[\tilde{\nu}(t) = \tilde{\nu}(s)] = e^{-W(s, t)}$ を得てポワソン過程を特別な場合として得る.

定理 1 に戻って, 確率順位付けモデルの定式化に出てこない記号は位置強度結合分布の極限 μ_t だけなので, この決定が定理の根幹である. 強度 w_i が定数や t だけの関数の場合は $\tilde{\nu}_i^{(N)}$ は独立増分で i について独立であり, その結果ポワソン分布に由来する指数関数 (2) を用いて (13) のように μ_t を書けたが, w_i に位置 x 依存性があると (14) のように x に従属確率変数である $Y_j^{(N)}$ たちを代入するので, 従属性の影響は極限でも非自明に残る. そこで, 初期点 $(y_0, 0)$ と境界点 $(0, t_0)$ を合わせて γ と置いて γ と時刻 t の関数 $\theta(\gamma, t)$ (であって (6) のように粒子系の軌道の極限になるようなもの) を (14) の $Y_j^{(N)}$ の代わりに強度 $w \in W$ の位置変数に代入したものを $\tilde{\nu}_{\theta, w, z}$ とおくと, i についての従属性が消えることに注目する. $\tilde{w}_{\theta, w, z}(s, t) = \begin{cases} w(\theta((z, 0), t), t), & s = 0, \\ w(\theta((0, s), t), t), & s > 0, \end{cases}$ と置くと, $\tilde{\nu}_{\theta, w, z}$ は $\tilde{w}_{\theta, w, z}$ を強度とする, 直前の到着時刻に強度が依存する点過程である. この点

過程に基づく確率順位付けモデル (1) を流れ θ が定める強度に従う確率的順位付けモデルと呼ぶ [12] . μ_t は (12) の極限 φ を通して $\varphi_{y_C}(dw, (y_0, t_0), t) = \mu_t(dw \times [y_C((y_0, t_0), t), 1])$, から定まることは第 2 節と同様だが, 定理 1 では従属確率変数の和の極限なので独立確率変数列の和についての大量の完全法則では証明に足りない . 一方, 流れが定める強度に従う確率的順位付けモデルは個々の粒子の先頭に跳ぶ時刻は独立増分性がないが, 粒子間は独立なので大量の完全法則が使えて,

$$\varphi_\theta(dw, \gamma, t) = \int_{z \in [y_0, 1]} P[\tilde{v}_{\theta, w, z}(t) = \tilde{v}_{\theta, w, z}(t_0)] \mu_0(dw \times dz), \quad (15)$$

が成り立つ . 最後に, θ は $\mathcal{G}(\theta)(\gamma, t) = 1 - \varphi_\theta(W, \gamma, t)$ で定まる流れの集合上の写像の固定点に選ぶと, $N \rightarrow \infty$ で「中間モデル」である流れが定める強度に従う確率的順位付けモデルと本来の確率的順位付けモデルの極限が一致することが Gronwall 不等式型の議論によって証明できる [12] .

この, \mathcal{G} の固定点ということを用微分方程式の言葉で言い換えると「蒸発だけで 1 方向に流れる 1 次元多成分流体を表す偏微分方程式系の蒸発率 w が位置に依存する場合の非局所項 (積分項) を持つ一般化」の解を直前の到着時刻に強度が依存する点過程を用いて書ける, と理解できる . 偏微分方程式との関係を示すために, 位置 y と時刻 t と流体成分 α に依存する蒸発率 w の分布が $\lambda = \sum_{\beta} r_{\beta} \delta_{w_{\beta}}$ の場合に, 流体成分 α の密度

$\frac{\partial \mu_t(\{w_{\alpha}\} \times dy)}{\partial y}$ の下流側の積分 $U_{\alpha}(y, t) = \mu_t(\{w_{\alpha}\} \times [y, 1])$ が満たす方程式を書く . まず, w が位置に依存しない場合は,

$$\frac{\partial U_{\alpha}}{\partial t}(y, t) + \sum_{\beta} w_{\beta}(t) U_{\beta}(y, t) \frac{\partial U_{\alpha}}{\partial y}(y, t) = -w_{\alpha}(t) U_{\alpha}(y, t), \quad (y, t) \in [0, 1] \times [0, T]. \quad (16)$$

となる . たとえば, w が定数の場合の (13) は (微分 t'_C の表式を用いると) (16) を満たすことを確かめられる . (16) が特性曲線 $\frac{dy_C}{dt}(t) = \sum_{\beta} w_{\beta}(t) U_{\beta}(y_C(t))$ を用いて解けることは古典 [2] だが, w が位置 y にも依存する場合への偏微分方程式系 (16) と特性曲線の自然な一般化は, 非局所項を持つ

$$\frac{\partial U_{\alpha}}{\partial t}(y, t) - \sum_{\beta} \int_y^1 w_{\beta}(z, t) \frac{\partial U_{\beta}}{\partial z}(z, t) dz \frac{\partial U_{\alpha}}{\partial y}(y, t) = \int_y^1 w_{\alpha}(z, t) \frac{\partial U_{\alpha}}{\partial z}(z, t) dz,$$

および, $\frac{dy_C}{dt}(t) = - \sum_{\beta} \int_{y_C(t)}^1 w_{\beta}(z, t) \frac{\partial U_{\beta}}{\partial z}(z, t) dz$, であることがわかる [9, §1] . この解は (12) などから推測できるとおり) $U_{\alpha}(y_C(t), t) = \varphi_{y_C}(\{w_{\alpha}\}, \gamma, t)$ と, (15) によって直前の到着時刻に強度が依存する点過程で表せる . 古典的な偏微分方程式の自然な一般化が (見つけにくい) 非局所項を持ち, ポワソン過程の非独立増分な一般化であらわに解ける .

Amazon ランキングと先頭に跳ぶ規則を結びつけた背景や直感等については [7] を, 本稿後半の数学的な内容についてのもう少し詳しい紹介は [14] を, 参照いただければなおさいわいである . 本稿では強度の時刻依存性と社会行動の日周期の検討も省略したがこれらについても [7, 14] にもう少し詳しく紹介した . <http://web.econ.keio.ac.jp/staff/hattori/amazonj.htm#top> 以下にも記事を貼ってある .

参考文献

- [1] C. Anderson, *The Long Tail: Why the Future of Business Is Selling Less of More*, Hyperion Books, 2006.
- [2] A. Bressan, *Hyperbolic systems of conservation laws, The one-dimensional Cauchy problem*, Oxford Univ. Press, Oxford, 2005.
- [3] Y. Hariya, K. Hattori, T. Hattori, Y. Nagahata, Y. Takeshima, T. Kobayashi, *Stochastic ranking process with time dependent intensities*, *Tohoku Mathematical Journal* **63** (1) (2011) 77–111.
- [4] K. Hattori and T. Hattori, *Existence of an infinite particle limit of stochastic ranking process*, *Stochastic Processes and their Applications* **119** (2009) 966–979.
- [5] K. Hattori and T. Hattori, *Equation of motion for incompressible mixed fluid driven by evaporation and its application to online rankings*, *Funkcialaj Ekvacioj* **52** (2009) 301–319.
- [6] K. Hattori and T. Hattori, *Sales ranks, Burgers-like equations, and least-recently-used caching*, *RIMS Kokyuroku Bessatsu* **B21** (2010) 149–162.
- [7] 服部哲弥「Amazon ランキングの謎を解く - 確率的な順位付けが教える売上の構造」, 化学同人出版, 2011 .
<http://web.econ.keio.ac.jp/staff/hattori/amazonj.htm>
<http://web.econ.keio.ac.jp/staff/hattori/amazone.htm>
- [8] T. Hattori, *Stochastic ranking process and web ranking numbers*, in *Mathematical Quantum Field Theory and Renormalization Theory*, T. Hara, T. Matsui, F. Hiroshima, eds., *Math-for-Industry Lecture Note Series* **30** (2011) 178–191.
- [9] T. Hattori, *Point process with last-arrival-time dependent intensity and 1-dimensional incompressible fluid system with evaporation*, *Funkcialaj Ekvacioj* **60** (2017) 171–212.
- [10] T. Hattori, *Open problems to an infinite system of quasi-linear partial differential equations with non-local terms*, *Symposium on Probability Theory 2014/12/16–19*, *RIMS Kokyuroku* **1952** (2015) 9–16.
- [11] T. Hattori, *Doubly uniform complete law of large numbers for independent point processes*, *Journal of Mathematical Sciences the University of Tokyo* **25** (2018) 171–192.
- [12] T. Hattori, *Cancellation of fluctuation in stochastic ranking process with space-time dependent intensities*, *Tohoku Mathematical Journal* **71**(3) (2019) 掲載予定 . <http://arxiv.org/abs/1612.09398>.
- [13] 服部哲弥「確率変数の収束と大数の完全法則 - 少しマニアックな確率論入門」, 共立出版, 2019 .
- [14] 服部哲弥「確率的順位付けの数理モデル」*数学* **71** (2019) 掲載予定 .
- [15] T. Hattori, S. Kusuoka, *Stochastic ranking process with space-time dependent intensities*, *Latin American Journal of Probability and Mathematical Statistics* **9** (2) (2012) 571–607. *ALEA, Lat. Am. J. Probab. Math. Stat.* **9** (2) (2012) 571–607.
- [16] Y. Nagahata, *Tagged particle dynamics in stochastic ranking process*, *Kodai Mathematical Journal* **36** (3) (2013) 397–408.
- [17] Y. Nagahata, *Functional central limit theorem for tagged particle dynamics in stochastic ranking process*, *Kodai Mathematical Journal* **36** (3) (2013) 409–427.
- [18] M. L. Tsetlin, *Finite automata and models of simple forms of behaviour*, *Russian Mathematical Surveys* **18** (1963) 1–27.