

# Amazon.co.jp のランキングを記述する偏微分方程式

服部哲弥 (東北大学・理)  
服部久美子 (首都大学東京・数学)

蒸発を動因とする 1 次元非圧縮性混合流体の非線形偏微分方程式の解を与える。この解はランキングを記述する確率モデル (stochastic ranking model) の無限粒子極限として得られることが分かっている。このことから、この偏微分方程式とその解は 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しうる。このことを実際の web データで実証し、結果の経済学的意味にも言及する。

## 1. 蒸発を動因とする 1 次元非圧縮性混合流体の運動

有限区間  $y \in [0, 1)$  上の複数種類  $i = 1, 2, \dots$  の成分からなる混合流体を考える。各成分  $i$  に対して正定数  $f_i > 0$  (蒸発率) が与えられている。位置  $y$  時刻  $t$  での成分  $i$  の密度 (成分比)  $u_i(y, t)$  の時間発展を記述する以下の偏微分方程式系の初期値問題を考える:  $t \geq 0, 0 \leq y < 1$  に対して

$$\begin{aligned} \frac{\partial u_i(y, t)}{\partial t} + \frac{\partial (v(y, t) u_i(y, t))}{\partial y} &= -f_i u_i(y, t), \quad i = 1, 2, \dots, \\ \frac{\partial v(y, t)}{\partial y} &= -\sum_j f_j u_j(y, t), \\ \sum_j u_j(y, t) &= 1, \end{aligned}$$

初期値:  $u_i(y, 0) \geq 0, 0 \leq y < 1, i = 1, 2, \dots$ ,

境界条件: 初期値の空間積分 ( $i$  成分の総量) を  $\rho_i = \int_0^1 u_i(y, 0) dy$  とおくと、 $t \geq 0$  に対して

$$\begin{aligned} v(1 - 0, t) &= 0, \\ u_i(0, t) &= \frac{f_i \rho_i}{\sum_j f_j \rho_j} > 0, \quad i = 1, 2, \dots \end{aligned}$$

この方程式系は、 $f_i$  を蒸発率とする蒸発を伴う非圧縮性の混合流体の密度  $u_i(y, t)$  の時間発展を記述し、非線形性をもたらす速度場は、流れが蒸発分を埋めることだけで生じることを表す。応用上は非負値解 ( $u_i(y, t) \geq 0$ ) に興味がある。

上流側 ( $y = 0$ ) の境界条件が初期値によって tune されているが、これは (残りの方程式の下で) 各成分の総量について保存系であること

$$\int_0^1 u_i(y, t) dy = \rho_i, \quad t \geq 0,$$

と同値になることが分かる。すなわち、 $0 \leq y < 1$  の各点で蒸発した総量が即座に上流側から凝結・流入して全体として保存系になるように境界条件を定めた (下流側 ( $y = 1$ ) は速度 0、すなわち物質の出入りのない壁。)

以下の結論は 2 種以上無限種まで任意の種類数で成立する。さらに、蒸発率として連続値を許して ( $f_i \Rightarrow w$ )、 $u_i$  をその上の確率測度と置き換え ( $u_i \Rightarrow \mu(dw)$ ) ても、以下の結論は自然な拡張の意味で成り立つ。Web 上のデータへの応用においては、非常に大きな種類数  $N$  を考えて極限をとって連続無限種類の言葉で書くのが分かりやすい。

## 2. 初期値問題の解

考えている方程式系の流体方程式としてのユニークな点は、

$$y_C(t) = 1 - \sum_i \rho_i e^{-f_i t}$$

なる点を境にして、その左右で解の性質が全く異なることである。考えている方程式が空間1次元で流体粒子が渦などで混ざることが無く、「ところてん」式に右に押し流され、蒸発した分は左端からまとめて供給されるためである。 $y_C(t)$  は  $y = 0$  にあった流体粒子が（蒸発しないとき） $t$  時間後に達する位置を表す。 $y_C$  は狭義減少なので逆関数  $t_0: [0, 1) \rightarrow [0, \infty)$  が存在することに注意する。 $t_0(y)$  は  $y = 0$  にあった流体粒子が位置  $y$  に達するまでの時間を表す。

次に示す解は、 $y < y_C(t)$  では定常 ( $\frac{\partial u_i}{\partial t}(y, t) = 0$ ) であり、 $y > y_C(t)$  では波動の伝搬のように初期値の情報が（蒸発分を除いて）そのまま残ることが特徴である。命題。以下の  $u_i$  は考えている方程式系の初期値問題の解である。

$$u_i(y, t) = \begin{cases} \frac{f_i e^{-f_i t_0(y)} \rho_i}{\sum_j f_j e^{-f_j t_0(y)} \rho_j}, & y < y_C(t), \\ \frac{e^{-f_i t} u_i(\hat{y}(y, t), 0)}{\sum_j e^{-f_j t} u_j(\hat{y}(y, t), 0)}, & y > y_C(t). \end{cases}$$

ここで、下流側  $y > y_C(t)$  の解に出てくる  $\hat{y}$  は各時刻  $t \geq 0$  ごとに  $y$  の関数  $\hat{y}(\cdot, t): [y_C(t), 1) \rightarrow [0, 1)$  とみて、時刻  $t$  に位置  $y$  にいる流体粒子の  $t = 0$  での初期位置を表す。より具体的には、 $y_C(t)$  を拡張して、 $0 \leq y < 1$  に対して、初期位置  $y$  の流体粒子の時刻  $t$  での途中で蒸発しなかった場合の位置は

$$y_C(y, t) = 1 - \sum_j e^{-f_j t} \int_y^1 u_j(z, 0) dz$$

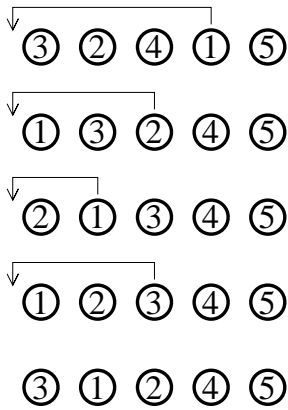
で与えられる ( $y_C(t) = y_C(0, t)$ ) が、その  $y$  の関数としての逆関数が  $\hat{y}$  である。◇

$\lim_{t \rightarrow \infty} y_C(t) = 1$  なので、初期位置に依存する部分の幅が狭まる形の緩和が起きている。

## 3. Stochastic ranking process の流体力学極限

考えている非線形偏微分方程式は、物理的意味が明確で、解もあらわに分かる方程式であるにもかかわらず、注目されてきた気配が無い。現実の流体の場合、蒸発分の補給で生じる流れは弱すぎて表面張力や重力の効果に隠れるからであろう（たとえば赤穂の製塩でも樋を傾けて重力によって流しているようだ。）また2次元以上の流れでは渦が生じて「ところてん」式の流れは乱されるので空間を高次元化すると解をあらわに見つけるのは難しくなるだろう。

現実の流体としては以上のように応用が限られているために注目されなかったのかもしれないが、興味深いことに、Amazon.co.jp などのランキングや 2ch.net などの掲示板のスレッド一覧にこの方程式が記述しうる現象が実在する。両者を数学的に結びつけるのは stochastic ranking process と呼ぶ確率モデルである。



$N$  個の粒子を 1 位から  $N$  位まで一列に並べる．それぞれの粒子を区別するために  $i = 1, 2, \dots, N$  なる番号が前もって各粒子に記されているとする（図の中の数字は順位ではなく粒子を区別する番号）各粒子  $i$  には跳躍率  $f_i > 0$  が定まっていて，短い時間  $\Delta t$  時間に 1 位に飛ぶ確率が常に  $f_i \Delta t + O(\Delta t^2)$  であるとする（言い換えると， $t$  時間 jump が起きない確率が指数分布  $e^{-f_i t}$  で与えられるとする）．たとえば時刻  $t$  に  $r$  位にいた粒子が 1 位に飛んだとすると，直前まで 1 位から  $r - 1$  位にいた粒子はそれぞれ順送りに 2 位から  $r$  位に位置を変える．粒子  $i$  の時刻  $t$  での（ランダムな）位置を  $X_i(t)$  と書き，その初期位置  $X_i(0) = x_i$  は定まっているものとする．

直感的には，時間とともに列（順位）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な従属確率変数に関する大数の法則を示すことによって，経験分布の極限としてこの直感を数学的に厳密かつ精密に記述できる．結論は， $i = 1, 2, \dots$  に対して，蒸発率が  $f_i$  に等しい粒子についての空間的にスケールした初期順位  $y_{i'} = \frac{1}{N}(x_{i'} - 1)$  の分

布  $\frac{1}{N} \sum_{i'; f_{i'}=f_i} \delta(y - y_{i',0})$  がなめらかな密度  $u_i(y, 0) \geq 0$  を持つ分布に「近い」とき

$(\sum_i u_i(y, 0) = 1)$ ，jump 率  $f_i$  の粒子の総密度を  $\rho_i = \int_0^1 u_i(z, 0) dz$  とおくと，時刻  $t$

での空間的にスケールされた位置  $Y_i(t) = \frac{1}{N}(X_i(t) - 1)$  の経験分布（分布値確率変

数） $\frac{1}{N} \sum_{i'; f_{i'}=f_i} \delta(y - Y_{i'}(t))$  は密度  $u_i(y, t)$  を持つ分布に「近い」．ここで  $u_i(y, t)$  は

2 節で書いた偏微分方程式の解  $u_i(y, t)$  に等しい（より一般的な記述と「近い」という言葉（経験分布の  $N \rightarrow \infty$  での収束）の正確な内容は文献の定理を参照していただきたい）

要約すると，stochastic ranking process という簡単な多粒子系の確率モデルの「流体力学極限」として考えている非線形偏微分方程式を得る．個々の粒子の 1 位（最上流）への jump はランダムだが，上位（上流）側は jump 率の高い粒子が多く下位（下流）側は率の低い粒子が時間とともに増えていく，という空間密度に関する性質は  $N$  が大きいとき決定論的である．また，各粒子の 1 位になってからの順位の時間変化  $x_C(t)$  も，粒子が流れに乗っている間は  $N$  が大きいとき  $N y_C(t) + 1$  に近い．

#### 4 . Amazon.co.jp のランキング

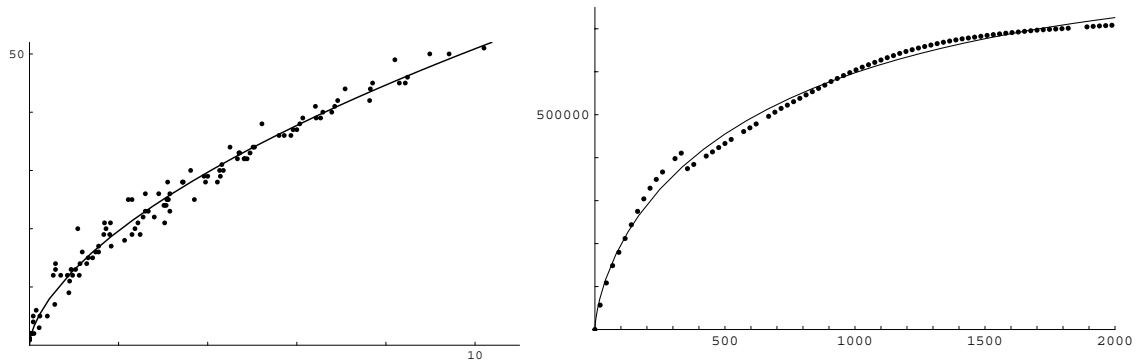
以上の理論を実際の状況に応用するためには，蒸発率の分布  $\{(\rho_i, f_i)\}$  を定める必要がある．社会学や経済学では Pareto 分布

$$f_i = a \left( \frac{N}{i} \right)^{1/b}, \quad \rho_i = 1, \quad i = 1, 2, \dots, N,$$

（統計学では log-linear 分布，物理学では power law と呼ばれることも多い）が使われることが多い．ここで  $a, b > 0$  は定数で，たとえば所得の議論では  $a$  は最低収入を表し，指数  $b$  は所得格差（ $b$  小  $\Leftrightarrow$  不平等， $b$  大  $\Leftrightarrow$  平等）を表す．Pareto 分布を用いると， $N$  が大きいとき，時刻 0 に 1 位にいた粒子の時刻  $t$  での順位は

$$x_C(t) = N y_C(t) + 1 \simeq N - Nb(at)^b \Gamma(-b, at)$$

となり,  $N, a, b$  を与えれば軌道が決まる.  $\Gamma(z, p) = \int_p^\infty e^{-w} w^{z-1} dw$  は不完全ガンマ関数. この  $x_C(t)$  を 2ch.net のスレッド一覧の順位の変化と Amazon.co.jp の本の



ランキングの実測データに当てはめた例がそれぞれ左図と右図である. いずれも横軸は時間, 縦軸は順位, 点が実測値, 曲線が  $x_C(t)$ . Pareto 分布のパラメータ  $a, b$  (Amazon.co.jp では  $N$  も) は最小 2 乗法でデータから定めた. この図を見ると, 我々の単純な数学モデルは, その単純さに比して実際のデータをよく説明していると感じる.

特に指数  $b$  の最尤値は 2ch.net と Amazon.co.jp のいずれでも  $b \simeq 0.6$  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' (順位の下位側) を過大評価することになる.

我々の方法は応用上重要な指数  $b$  を順位の時間変化のデータから直接得ることができる. その背景にはランキング上の 1 粒子の軌道によって多数の粒子 (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数学的裏付け (大数の法則) がある. このような試みはこれまで無かったようであり, 研究に値すると考える.

#### 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.