Equation of motion for incompressible mixed fluid driven by evaporation and its application to online rankings

Kumiko Hattori
Department of Mathematics and Information Sciences, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan.
email: khattori@tmu.ac.jp

Tetsuya Hattori
Mathematical Institute, Graduate School of Science, Tohoku University, Sendai 980-8578, Japan.
URL: http://www.math.tohoku.ac.jp/hattori/amazone.htm
email: hattori@math.tohoku.ac.jp

October 10, 2008

ABSTRACT

We consider a system of partial differential equations for the densities of components of one dimensional incompressible fluid mixture whose motion is driven by evaporation. We prove existence and give explicit form of unique global classical non-negative solution to initial value problem for the system. The solution is known to be an infinite particle limit of stochastic ranking processes, which is a simple stochastic model of time evolutions of Amazon Sales Ranks. As a practical application, we collected data from the web and performed statistical fits to our formula. The results suggest that the system of equations and solutions, though very simple, may have applications in the analysis of online rankings.

Keywords: evaporation driven fluid; non-linear wave; stochastic ranking process; long tail; online ranking;

2000 Mathematics Subject Classification: Primary 35C05; Secondary 35Q35, 82C22

Corresponding author: Tetsuya Hattori, hattori@math.tohoku.ac.jp
Mathematical Institute, Graduate School of Science, Tohoku University, Sendai 980-8578, Japan
tel+FAX: 011-81-22-795-6391

1 Introduction.

Let $f_i \geqq 0$, $i = 1, 2, \cdots$, be non-negative constants, and consider the following system of non-linear partial differential equations for the functions $u_i(y,t)$, $i = 1, 2, \cdots$, and $v(y,t)$, defined on $(y,t) \in [0,1) \times \mathbb{R}_+$:

$$\frac{\partial\, u_i(y,t)}{\partial t} + \frac{\partial\, (v(y,t)\, u_i(y,t))}{\partial y} = -f_i u_i(y,t), \quad i = 1, 2, \cdots, \tag{1}$$

$$\sum_j u_j(y,t) = 1\,. \tag{2}$$

We consider initial value problems for smooth non-negative initial data $u_i(y,0) = u_i(y)$;

$$u_i(y) \geqq 0,\ i = 1, 2, \cdots,\ \sum_j u_j(y) = 1, \quad 0 \leqq y < 1,$$

with the boundary conditions at $y = 0$ and $y = 1$:

$$v(1-0,t) = 0, \tag{3}$$

$$u_i(0,t) = \frac{f_i \rho_i}{\displaystyle\sum_j f_j \rho_j}, \quad i = 1, 2, \cdots, \tag{4}$$

for $t \geqq 0$, where, for each i,

$$\rho_i = \int_0^1 u_i(z)\, dz, \tag{5}$$

and we assume

$$0 < \sum_j f_j \rho_j < \infty. \tag{6}$$

Note that adding up (1) over i and applying (2) we have

$$\frac{\partial\, v(y,t)}{\partial y} = -\sum_j f_j u_j(y,t), \tag{7}$$

which, with (3), determines v in terms of u_i. Note also that (2) obviously implies

$$\sum_j \rho_j = 1. \tag{8}$$

Given the constants f_i and the initial data $u_i(y)$, the set of equations (1) (2) (3) (4) defines the evolution of our system. The following arguments and results hold both for finite components ($i = 1, 2, \cdots, N$) and infinite components. (In fact, we can extend the system and the solution to a case with any probability space Ω, by replacing $u_i(y,t)$ with a measure $\mu(d\omega, y, t)$. See [4] for probability theoretic arguments.)

A physical meaning of the system is as follows. We are considering a motion of incompressible fluid mixture in an interval of length 1, where $u_i(y,t)$ is the density of i-th component at space-time point (y,t). (2) implies that we normalize total density of the fluid mixture to be 1 at each space-time point (y,t), hence $u_i(y,t)$ represents the ratio of i-th component at (y,t). We are naturally interested in the non-negative solutions $u_i(y,t) \geqq 0$. $v(y,t)$ is the velocity field of the fluid. (1) is the equation of continuity, and the right-hand-side implies that each component evaporates with rate f_i per unit time and unit mass. Note that the set of equations (7) and (3) is equivalent to

$$v(y,t) = \sum_j f_j \int_y^1 u_j(z,t)dz, \tag{9}$$

which implies that the velocity field, or the motion of the fluid, is caused solely by filling the amount of fluid which evaporated from the right side of the point y. In particular, we have no flux through the boundary $y = 1$ ($v(1-0,t) = 0$).

The boundary condition (4) at $y = 0$ is so tuned by the initial data that the loss of mass by evaporation is compensated by the immediate re-entrance at $y = 0$ as liquidized particles, so that the total mass of each fluid component in the interval $[0,1)$ is conserved over time:

$$\int_0^1 u_i(z,t)\,dz = \rho_i, \quad t \geqq 0. \tag{10}$$

In fact, (10) is equivalent to (4), under the conditions (1) (2) (3) (5), if $\sup_i f_i < \infty$. (See Appendix A.)

Let

$$y_C(y,t) = 1 - \sum_j e^{-f_j t} \int_y^1 u_j(z)\,dz, \quad 0 \leqq y < 1,\ t \geqq 0. \tag{11}$$

For each $t \geqq 0$, $y_C(\cdot,t)\ :\ [0,1) \to [y_C(0,t),1)$ is a continuous, strictly increasing, onto function of y, and its inverse function $\hat{y}(\cdot,t) :\ [y_C(0,t),1) \to [0,1)$ exists:

$$1 - y = \sum_j e^{-f_j t} \int_{\hat{y}(y,t)}^1 u_j(z)\,dz, \quad y_C(0,t) \leqq y < 1,\ t \geqq 0. \tag{12}$$

$y_C(y,t)$ denotes the position of a fluid particle at time t (on condition that it does not evaporate up to time t) whose initial position is y. $\hat{y}(y,t)$ denotes the initial position of a fluid particle located at y ($\geqq y_C(0,t)$) at time t.

With slight abuse of notations, we will often write $y_C(t)$ for $y_C(0,t)$:

$$y_C(t) = y_C(0,t) = 1 - \sum_j \rho_j e^{-f_j t}, \quad t \geqq 0. \tag{13}$$

Note that $\sum_j f_j \rho_j > 0$, as assumed in (6), implies that $y_C(t)$ is strictly increasing (as well as continuous) in t, and (8) implies $y_C(0) = 0$. Hence, for each $0 < y < y_C(t)$ there exists unique $t_0 = t_0(y) \in (0,t)$ such that

$$y_C(t_0(y)) = y, \quad 0 \leqq y < y_C(t),\ t > 0. \tag{14}$$

In Section 2 we prove the following.

Theorem 1 *There exists a unique global (i.e., for all $t \geqq 0$) non-negative classical solution to the initial value problem for the system of partial differential equations defined by (1)–(5), which is explicitly given by (9) and*

$$u_i(y,t) = \begin{cases} \dfrac{e^{-f_i t_0(y)} f_i \rho_i}{\sum_j e^{-f_j t_0(y)} f_j \rho_j}, & t > t_0(y), \\ \dfrac{e^{-f_i t} u_i(\hat{y}(y,t))}{\sum_j e^{-f_j t} u_j(\hat{y}(y,t))}, & 0 \leqq t < t_0(y), \end{cases} \qquad \diamond \tag{15}$$

Note that for $t > t_0(y)$ the solution is stationary:

$$\frac{\partial u_i}{\partial t}(y,t) = 0, \quad t > t_0(y). \tag{16}$$

For $t < t_0(y)$, effect of initial conditions exists in the form of wave propagation.

In natural phenomena where evaporation is active, such as producing salt out of sea water, viscosity, surface tension, and external forces such as gravitational forces dominate, and the effect of evaporation on the motion of fluid would be relatively too small to observe. Thus the equation and the solution we consider in this paper may not have attracted much practical attention. However, there are phenomena on the web for which our formulation may work as a simplified mathematical model, such as the time evolutions of rankings of book sales in the online booksellers. Such possibility is theoretically based on a result that (15) appears as an infinite particle limit of the stochastic ranking process [4], which is a simple model of the time evolutions of e.g., the number known as the Amazon Sales Rank. (We note that this number has mathematically little to do with the perhaps more popular notion of Google Page Ranks.) We collected data of the time evolution of the numbers from the web, and performed statistical fits of the data to (13). Considering the simplicity of our model and formula, we find the fits to be rather good. The results suggest that there is a new application of our results in the analysis of online rankings.

The plan of this paper is as follows. In Section 2 we give a proof of Theorem 1. In Section 3 we recall the stochastic ranking process defined in [4], and relate (15) to the infinite particle limit studied in [4]. In Section 4 we give results of fits to (13) of data from the web.

The authors would like to thank Prof. T. Miyakawa, Prof. M. Okada, Prof. H. Nawa, and Prof. T. Namiki for taking interest in our work and for inviting the authors to their seminars. The research of K. Hattori is supported in part by a Grant-in-Aid for Scientific Research (C) 16540101 from the Ministry of Education, Culture, Sports, Science and Technology, and the research of T. Hattori is supported in part by a Grant-in-Aid for Scientific Research (B) 17340022 from the Ministry of Education, Culture, Sports, Science and Technology.

2 Proof of the main theorem.

Put

$$U_i(y,t) = \int_y^1 u_i(z,t)\,dz, \quad i = 1, 2, \cdots, \tag{17}$$

and $U_i(y) = \int_y^1 u_i(z)\,dz$ for the initial data $U_i(y,0) = U_i(y)$. With (17), the the system of equations (1)–(5) is equivalent to the following: $U_i(y,t)$ is decreasing in y and $U_i(1-0,t) = 0$, and

$$\frac{\partial U_i}{\partial t}(y,t) + v(y,t)\,\frac{\partial U_i}{\partial y}(y,t) = -f_i U_i(y,t), \quad i = 1,2,\cdots, \tag{18}$$

$$\sum_j U_j(y,t) = 1 - y\,, \tag{19}$$

$$v(y,t) = \sum_j f_j\, U_j(y,t), \tag{20}$$

for $0 \leqq y < 1$, $t \geqq 0$ (note (9)), and noting (10),

$$U_i(0,t) = \rho_i \quad i = 1,2,\cdots,\ t \geqq 0. \tag{21}$$

The remainder of the proof is a simple application of characteristics. For each $0 \leqq y_0 < 1$ let $y_B = y_B(y_0;t)$ be a solution to an ODE

$$\frac{d\,y_B}{dt}(y_0;t) = v(y_B(y_0;t),t),\ t \geqq 0, \quad y_B(y_0;0) = y_0\,, \tag{22}$$

and put

$$\phi_i(t) = U_i(y_B(y_0;t),t). \tag{23}$$

With (22) and (18) it follows that

$$\begin{aligned}\frac{d\,\phi_i}{dt}(t) &= \frac{\partial U_i}{\partial t}(y_B(y_0;t),t) + v(y_B(y_0;t),t)\,\frac{\partial U_i}{\partial y}(y_B(y_0;t),t)\\ &= -f_i U_i(y_B(y_0;t),t) = -f_i\phi_i(t),\end{aligned}$$

hence, with $\phi_i(0) = U_i(y_0,0) = U_i(y_0)$, ϕ_i is uniquely solved as

$$\phi_i(t) = U_i(y_0)\,e^{-f_i t}. \tag{24}$$

With (20) and (22), we then find

$$\frac{d\,y_B}{dt}(y_0;t) = \sum_j f_j U_j(y_0)\,e^{-f_j t},$$

hence, using (19) and (17), we have

$$y_B(y_0;t) = 1 - \sum_j U_j(y_0)\,e^{-f_j t} = 1 - \sum_j \int_{y_0}^1 u_j(z)\,dz\,e^{-f_j t} = y_C(y_0,t), \tag{25}$$

where y_C is defined in (11). With (12), (23) and (24) we uniquely obtain

$$U_i(y,t) = U_i(\hat{y}(y,t))\,e^{-f_i t}.$$

Differentiating by y and using (17) and (25)

$$u_i(y,t) = \left(\frac{\partial\, y_B}{\partial y_0}(\hat{y}(y,t);0)\right)^{-1} u_i(\hat{y}(y,t))\, e^{-f_i t} = \frac{u_i(\hat{y}(y,t))\, e^{-f_i t}}{\sum_j u_j(\hat{y}(y,t))\, e^{-f_j t}}\,,$$

where $\frac{\partial\, y_B}{\partial y_0}$ is the derivative of $y_B = y_B(y_0;t)$ with respect to the parameter y_0. This proves (15) for $y > y_C(t)$, i.e., for $0 \leqq t < t_0(y)$.

Next let $t > t_0(y)$ and put $t_1 = t - t_0(y) \in (0,t)$. Let y_A be a solution to an ODE

$$y_A'(s) = v(y_A(s),s),\ \ s \geqq t_1\,,\quad y_A(t_1) = 0\,, \tag{26}$$

and put

$$\phi_i(s) = U_i(y_A(s),s),\ \ s \geqq t_1\,. \tag{27}$$

Note that (21) implies $\phi_i(t_1) = U_i(0,t_1) = \rho_i$, hence, as below (22), ϕ_i is uniquely solved as

$$\phi_i(s) = \rho_i\, e^{-f_i(s-t_1)}. \tag{28}$$

With (20) and (26), we then find

$$y_A'(s) = \sum_j f_j \rho_j\, e^{-f_j(s-t_1)},$$

Note that (5) and (2) imply $\sum_j \rho_j = 1$. Hence,

$$y_A(s) = 1 - \sum_j \rho_j\, e^{-f_j(s-t_1)} = y_C(s-t_1),\ \ s \geqq t_1\,. \tag{29}$$

where y_C is defined in (13).

Putting $s = t$ in (27) and (28), using (29), and recalling that $t_1 = t - t_0(y)$, we have, with (14),

$$U_i(y,t) = U_i(y_C(t_0(y)),t) = \rho_i\, e^{-f_i t_0(y)}.$$

Differentiating by y and using (17), (14) and (13),

$$u_i(y,t) = \left(\frac{d\, y_C}{dt}(t_0(y))\right)^{-1} \rho_i f_i\, e^{-f_i t_0(y)} = \frac{\rho_i f_i\, e^{-f_i t_0(y)}}{\sum_j \rho_j f_j e^{-f_j t_0(y)}}\,,$$

which proves (15) for $t > t_0(y)$. This completes a proof of Theorem 1.

Remark. Observing (18) as a Burgers type system, absence of shock is essential in our proof for unique existence of global classical solution and preservation of regularity. A sufficient condition for the global absence of shock is (with other assumed conditions such as stationary boundary conditions and non-negativity of U_js)

$$\sum_j \frac{\partial U_j}{\partial y}(y_0) \geqq -1, \tag{30}$$

which is satisfied by (19), or the incompressibility condition (2). In fact, without (19), we have, in place of (25),

$$y_B(y_0;t) = y_0 + \sum_j U_j(y_0)\,(1 - e^{-f_j t}),$$

which is strictly increasing in y_0 for any $t \geqq 0$, if (30) holds. Hence $\hat{y}(\cdot, t)$, the inverse function of $y_B(\cdot;t)$ exists, and consequently, $U_i(y,t)$ is solved. ◇

Before closing this section, we give a couple of examples of the solution to the equation of motion.

Example 1 (One particle type). For a pure fluid, $\rho_1 = 1$, hence we have $u_1(y,t) \equiv 1$. Since there is only one type of incompressible fluid, the density is constant. However, there is a flow driven by evaporation even in this case, and we actually have $y_C(t) = 1 - e^{-f_1 t}$.

Example 2 (Two particle types with $f_2 = 0$). Consider a mixture of 2 components with the ratios satisfying $\rho_1 > 0$ and $\rho_2 = 1 - \rho_1 > 0$. $f_2 = 0$ means no evaporation, so the situation is a salty sea with salt density ρ_2, where evaporation and flow from a river balance. We have

$$y_C(t) = \rho_1(1 - e^{-f_1 t}), \quad t_0(y) = -\frac{1}{f_1}\log(1 - \frac{y}{\rho_1}).$$

The expressions of $u_i(y,t)$ for $y < y_C(t)$ become simple:

$$v(y,t) = f_1\,(\rho_1 - y), \quad u_1(y,t) \equiv 1, \quad u_2(y,t) \equiv 0\,.$$

The pure water from the river comes in up to $y < y_C(t)$. (Note that we are considering a fictitious 1-dimensional case where no spacial mixing such as turbulence occurs and we have no other dynamics such as diffusion.)

If, furthermore, the initial distribution is uniform on $[0,1)$: $u_i(y) = \rho_i$, $i = 1,2$, then the expressions for $y > y_C(t)$ are also simple:

$$y_C(y,t) = 1 - (1-y)(\rho_1 e^{-f_1 t} + \rho_2), \quad \hat{y}(y,t) = 1 - \frac{1-y}{\rho_1 e^{-f_1 t} + \rho_2}\,,$$

and consequently,

$$v(y,t) = (1-y) f_1 \frac{\rho_1 e^{-f_1 t}}{\rho_1 e^{-f_1 t} + \rho_2}, \quad u_1(y,t) = \frac{\rho_1 e^{-f_1 t}}{\rho_1 e^{-f_1 t} + \rho_2}\,, \quad u_2(y,t) = \frac{\rho_2}{\rho_1 e^{-f_1 t} + \rho_2}\,,$$

for $y > y_C(t)$. (In general, the formulas are dependent on initial data in complex ways, and we do not have explicit formula.)

3 Infinite particle limit of stochastic ranking process.

In this section, we first recall the stochastic ranking process defined in [4], and relate its infinite particle limit proved in [4] to (15).

Let $(\Omega, \mathcal{B}, \mathrm{P})$ be a probability space. A stochastic ranking process of N particles, $\{X_i^{(N)}(t) \mid t \geqq 0,\ i = 1, 2, \cdots, N\}$, is defined as follows. We assume that the initial configuration $X_i^{(N)}(0) = x_{i,0}^{(N)}$, $i = 1, 2, \cdots, N$, is given, such that $x_{1,0}^{(N)}, x_{2,0}^{(N)}, \cdots, x_{N,0}^{(N)}$ is a permutation of $1, 2, \cdots, N$. For each i let $\tau_{i,j}^{(N)}$, $j = 0, 1, 2, \cdots$, be an increasing sequence of random jump times, such that $\{\tau_{i,j}^{(N)} \mid j = 0, 1, 2, \cdots\}$, $i = 1, 2, \cdots, N$, are independent (independence among particles), $\tau_{i,0}^{(N)} = 0$ and $\{\tau_{i,j+1}^{(N)} - \tau_{i,j}^{(N)} \mid j = 0, 1, 2, \cdots\}$ are i.i.d. with the law of $\tau_i^{(N)} = \tau_{i,1}^{(N)}$ being

$$\mathrm{P}[\ \tau_i^{(N)} \leqq t\ ] = 1 - e^{-w_i^{(N)} t},\ \ t \geqq 0, \tag{31}$$

where, for each i, $w_i^{(N)} > 0$ is a given constant (jump rate). (Note that with probability 1, $\tau_{i,j}^{(N)}$, $j = 0, 1, 2, \cdots$, is strictly increasing, and that $\tau_{i,j}^{(N)} \neq \tau_{i',j'}^{(N)}$ for any different pair of suffices $(i, j) \neq (i', j')$.) For each $i = 1, 2, \cdots, N$ we define the time evolution of $X_i^{(N)}$ by,

$$X_i^{(N)}(t) = x_{i,0}^{(N)} + \sharp\{i' \in \{1, 2, \cdots, N\} \mid x_{i',0}^{(N)} > x_{i,0}^{(N)},\ \tau_{i',1}^{(N)} \leqq t\},\quad 0 \leqq t < \tau_{i,1}^{(N)},$$

where $\sharp A$ denotes the number of elements in the set A, with $\sharp\emptyset = 0$, and for each $j = 1, 2, 3, \cdots$

$$\begin{aligned}&X_i^{(N)}(\tau_{i,j}^{(N)}) = 1, \quad \text{and}\\ &X_i^{(N)}(t) = 1 + \sharp\{i' \in \{1, 2, \cdots, N\} \mid \exists j' \in \mathbb{Z}_+;\ \tau_{i,j}^{(N)} < \tau_{i',j'}^{(N)} \leqq t\},\quad \tau_{i,j}^{(N)} < t < \tau_{i,j+1}^{(N)}.\end{aligned}$$

Proposition 2 ([4]) *Assume that the empirical distribution of jump rates converges weakly to a probability distribution* λ*:* $\lambda^{(N)}(dw) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \delta(w - w_i^{(N)})\, dw \to \lambda(dw),\ N \to \infty$. *Then*

$$\lim_{N\to\infty} \frac{1}{N} \sharp\{i \mid \tau_i^{(N)} \leqq t\} = y_C(t), \quad \textit{in probability,}$$

where

$$y_C(t) = 1 - \int_{[0,\infty)} e^{-wt} \lambda(dw), \quad t \geqq 0. \tag{32}$$

◇

Proposition 2 (and, the main theorem in [4]) holds for any distribution λ with finite mean (see [4] for details). Here in this section, in order to consider the situation consistent with that of Section 1, we will restrict ourself to the case where λ is a discrete measure, i.e., is concentrated on a finite or a countable set of non-negative numbers $\{f_1, f_2, f_3, \cdots\}$. Namely, we assume a form

$$\lambda(dw) = \sum_j \rho_j \delta(w - f_j) dw, \tag{33}$$

where $\delta(w - f_j)dw$ is a unit measure concentrated on f_j , and ρ_j are non-negative constants satisfying (8) and (6). Then (32) coincides with (13).

Let

$$Y_i^{(N)}(t) = \frac{1}{N}\,(X_i^{(N)}(t) - 1), \tag{34}$$

and

$$y_{i,0}^{(N)} = \frac{1}{N}\,(x_{i,0}^{(N)} - 1) \in [0,1) \cap N^{-1}\mathbb{Z}.$$

We consider the $N \to \infty$ limit of the empirical distribution on the product space of jump rate and spacial position;

$$\mu_t^{(N)}(dw, dy) = \frac{1}{N}\sum_i \delta(w - w_i^{(N)})\delta(y - Y_i^{(N)}(t))\, dw\, dy. \tag{35}$$

According to [4], we impose that the initial distribution converges weakly to a probability measure;

$$\mu_0^{(N)}(dw, dy) = \frac{1}{N}\sum_i \delta(w - w_i^{(N)})\delta(y - y_{i,0}^{(N)})\, dw\, dy \;\to\; \mu_0(dw, dy),\;\; N \to \infty \text{ (weakly.)} \tag{36}$$

To meet the notation in Section 1, we write

$$\mu_0(dw, dy) = \mu_{y,0}(dw)\, dy = \sum_j u_j(y)\, dy\, \delta(w - f_j)dw, \tag{37}$$

where we assume that $u_j(y)$ satisfies the assumptions in Section 1. Note that (5) and (2) respectively implies that the first and the second marginal of μ_0 is λ in (33) and the Lebesgue measure, respectively. Note also that (6) implies (with (33)) $\displaystyle\int w\lambda(dw) = \sum_j f_j\rho_j < \infty$. (In [4] it has also been assumed that $\lambda(\{0\}) = 0$ but the results hold in the present situation with (6).)

(37) implies

$$1 - \int_y^1 \int_0^\infty e^{-wt}\mu_0(dw, dy) = y_C(y, t),$$

where y_C is as in (11), hence its inverse function $\hat{y}(y, t)$ in y exists and is given by (12). Our notations are now consistent with those in [4], with the correspondence (33) and (37). The main theorem in [4] then implies that the joint empirical distribution of particle types and positions at time t

$$\mu_t^{(N)}(dw, dy) = \frac{1}{N}\sum_i \delta(w - w_i^{(N)})\delta(y - Y_i^{(N)}(t))\, dw\, dy \tag{38}$$

converges as $N \to \infty$ to a distribution $\mu_{y,t}(dw)\, dy$ on $\mathbb{R}_+ \times [0, 1)$. The limit distribution is absolutely continuous with respect to the Lebesgue measure on $[0, 1)$. The density $\mu_{y,t}(dw)$ with regard to y is given by

$$\mu_{y,t}(dw) = \begin{cases} \dfrac{we^{-wt_0(y)}\lambda(dw)}{\displaystyle\int_0^\infty we^{-wt_0(y)}\lambda(dw)}, & t > t_0(y), \\[2ex] \dfrac{e^{-wt}\mu_{\hat{y}(y,t),0}(dw)}{\displaystyle\int_0^\infty e^{-wt}\mu_{\hat{y}(y,t),0}(dw)}, & 0 \leqq t < t_0(y). \end{cases} \tag{39}$$

To see the correspondence between (39) and (15) in Section 2, let us rewrite (39) in terms of (33) and (37). For $y < y_C(t)$, we use (33) in (39) to find

$$\frac{we^{-wt_0(y)}\lambda(dw)}{\displaystyle\int_0^\infty we^{-wt_0(y)}\lambda(dw)} = \frac{\displaystyle\sum_j f_j e^{-f_j\, t_0(y)}\rho_j \delta(w-f_j)dw}{\displaystyle\sum_j f_j e^{-f_j t_0(y)}\rho_j}$$

which, according to (15), is equal to $\displaystyle\sum_j u_j(y,t)\delta(w-f_j)dw$. Similarly, for $y > y_C(t)$, we use (37) in (39) to find

$$\frac{e^{-wt}\mu_{\hat{y}(y,t),0}(dw)}{\displaystyle\int_0^\infty e^{-wt}\mu_{\hat{y}(y,t),0}(dw)} = \frac{\displaystyle\sum_j e^{-f_j t}u_j(\hat{y}(y,t))\,\delta(w-f_j)dw}{\displaystyle\sum_j e^{-f_j t}u_j(\hat{y}(y,t))} = \sum_j u_j(y,t)\,\delta(w-f_j)dw.$$

In both cases, we have the correspondence

$$\mu_{y,t}(dw) = \sum_j u_i(y,t)\,\delta(w-f_j)dw,$$

which relates the solution (15) of the PDE in Section 1 to the infinite particle limit of the stochastic ranking process in [4].

4 Possible application to rankings on the webs.

4.1 Pareto distribution.

The stochastic ranking process may be viewed as a mathematical model of the time evolution of rankings such as that of books on the online bookstores' web (e.g., www.Amazon.co.jp). In this example, N stands for the total number of books, i represents a specific title of a book, $w_i^{(N)}$ is the average rate with which the book i is sold, $x_{i,0}^{(N)}$ is the initial position (ranking) of the book, $\tau_{i,j}^{(N)}$ is the random time at which the book i is sold for the j-th time, and $X_i^{(N)}(t)$ is the ranking of the book i at time t.

A particle (a book, in the case of online booksellers) in the ranking jumps randomly to the rank 1 (each time a copy of the book is sold), and increases the ranking number by 1 each time some other particle of larger ranking number jumps to rank 1. Because an increase in ranking number is a result of jumps of very large number of particles in the tail side of the ranking, the particle effectively moves on the ranking queue in a deterministic way, even though each jump occurs at a random time. The jump corresponds to the evaporation in the fluid model. We can therefore predict the time evolutions of rankings appearing on the web rankings based on our model.

Here we try to see how a trajectory of a particle (13) could be observed in the web rankings. In applying (13) to the actual rankings, we need to choose the distribution of evaporation rates $\{(f_i, \rho_i) \mid i = 1, 2, \cdots, N\}$. In the case of social or economic studies such as online booksellers, this corresponds to choosing the distribution of activities or transactions. In the case of online booksellers, we have to choose the distribution of sales rates over books. The Pareto distribution (also called log-linear distribution in social studies, or power-law in physics literatures) is traditionally used as a basic model distribution for various social rankings, perhaps a most well-known example is the ranking of incomes. Let N be the total size of population, and for $i = 1, 2, \cdots, N$, denote by f_i the income of the i-th wealthiest person. If

$$f_i = a \left(\frac{N}{i} \right)^{1/b}, \ \rho_i = \frac{1}{N}, \quad i = 1, 2, 3, \cdots, N, \tag{40}$$

holds for some positive constants a and b, then the distribution of incomes is said to satisfy the Pareto distribution. (The Pareto distribution assumes all the constituents to have distinct f_i, which leads to equal weight $\rho_i = 1/N$ in our notation.) The factor a corresponds to the smallest income, and the exponent b reflects a social equality of incomes: in fact the ratio of the largest income to the smallest is $f_1/f_N = N^{1/b}$, which is close to 1 if b is large (a fair society), while is large (society is in monopoly) if b is small. (Our b corresponds to α in a standard textbook on statistics, θ in [3], and $-1/\beta_2$ in [2].)

Substituting (40) in (13) of Section 2, and approximating the summation by integration, we have, after a change of variable,

$$y_C(t) = 1 - b(at)^b \Gamma(-b, at) + O(N^{-1}), \tag{41}$$

where $\Gamma(z, p) = \displaystyle\int_p^\infty e^{-w} w^{z-1}\, dw$ is the incomplete Gamma function. $y_C(t)$ is a relative ranking normalized by N, so the time evolution of ranking $x_C(t)$ is

$$x_C(t) = 1 + N\, y_C(t). \tag{42}$$

The $O(N^{-1})$ contribution in (41) is (by a careful calculation) seen to be non-negative and bounded by $\dfrac{1}{N} e^{-at} \leqq \dfrac{1}{N}$, leading to a difference of at most 1 in the ranking $x_C(t)$, which is insignificant for our applications below, so we will ignore it.

Note that $\Gamma(-b, at) \to \infty$ as $t \to 0$, for $b > 0$. This divergence is harmless because it is cancelled by t^b in (41), but for numerical and asymptotic analysis, it is better to perform a partial integration on the right-hand side to find

$$y_C(t) = 1 - e^{-at} + (at)^b \Gamma(1 - b, at), \tag{43}$$

which, with (42), leads to

$$x_C(t) = N\left(1 - e^{-at} + (at)^b \Gamma(1 - b, at)\right) + 1. \tag{44}$$

The constant a, which denotes the lowest income in the Pareto distribution, has a role of a time constant in (44). In particular, the short time behavior of $x_C(t)$ for $0 < b < 1$ is

$$x_C(t) = c\, t^b + O(t), \tag{45}$$

where

$$c = Na^b\Gamma(1-b). \tag{46}$$

For $1 < b < 2$ we need a partial integration once more for a better expression;

$$y_C(t) = 1 - e^{-at}(1 - \frac{at}{b-1}) - \frac{(at)^b}{b-1}\,\Gamma(2-b,at) = \frac{ab}{b-1}\,t - \Gamma(2-b)\,\frac{a^b}{b-1}\,t^b + O(t^2). \tag{47}$$

Note that for $0 < b < 1$ the leading short time behavior is $y_C(t) = O(t^b)$, which is tangential to the y axis at $t = 0$, while for $b > 1$ (the case $b \geqq 2$ can be handled similarly) the linear dependence $y_C(t) = O(t)$ is dominant for small t.

4.2 Collected web bulletin board thread index listings.

2ch.net is one of the largest collected web bulletin boards in Japan. Each category ('board') has an index listing of the titles of 'threads' or the web pages in the board. The titles are ordered by "the last written thread at the top" principle; if one writes an article ('response') to a thread, the title of that thread in the index listing jumps to the top instantaneously, and the titles of other threads which were originally nearer to the top are pushed down by 1 in the listing accordingly. We can extract the exact time that a thread jumped to rank 1, because the time of each response in a thread is recorded together with the response itself. All these features of the 2ch.net index listing match the definition of the stochastic ranking process, hence 2ch.net is a suitable place for a testing the applicability of the stochastic ranking process.

We note that we can use deterministic (non-stochastic) formula such as (44) if N, the total number of threads in the board is large. In the case of 2ch.net, N is roughly about 700 to 800, so we would expect fluctuation of a few percent, and up to that accuracy, we expect a time evolution predicted by (44).

We also note that the time evolutions are independent of which thread one is looking at, because the changes in the ranking are caused by the collective motion of the threads towards the tail; popular threads jump back to the top ranking more frequently than the less popular ones, but as long as the threads remain in the queue (i.e., before the next jump), both a popular thread and an unpopular thread should behave in the same way, depending only on their position in the ranking.

We collected data of the time evolution of an index listing in the 2ch.net, and performed statistical fits of the data to (13). Fig. 1 is a plot of the threads in a board which jumped to the rank 1 during active hours one day and stayed in the queue without jumps until midnight. There were 12 such threads. The ranking is obviously monotone function of time between jumps, and there are no overtaking, so that the lines do not cross in the figure. Fig. 2 is a plot of same data as in Fig. 1, except that, for the horizontal axis the time is so shifted for each thread that the ranking of the thread is 1 at time 0. Though each thread starts at rank 1 on different time of the day, Fig. 2 shows that time evolutions after rank 1 are on a common curve. N is the total number of threads, which is $N = 795$ at the time of observation for Fig. 2, and a and b are positive constant parameters to be determined from the data. We performed a least square fit to (44) of $n_d = 117$ data points shown in Fig. 2. The best fit for the parameter set (a, b) is $(a^*, b^*) = (3.3425 \times 10^{-4}, 0.6145)$ ($\sqrt{\chi^2/n_d} \simeq 1.8$).

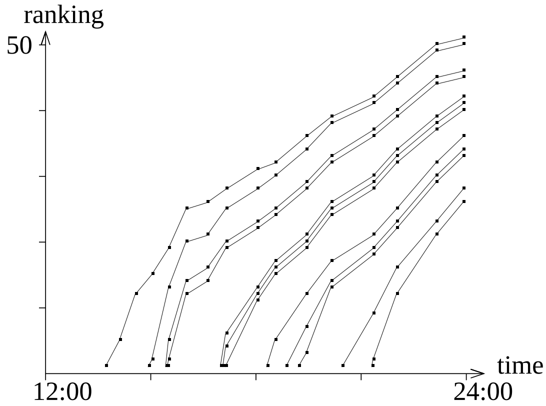

Fig 1: Record of ranking changes in an afternoon for 12 threads in a board of 2ch.net. Points from a thread are joined by line segments to guide the eye.

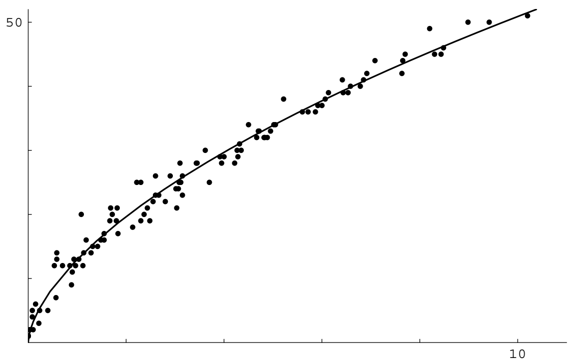

Fig 2: Collection of 12 threads in a board of 2ch.net, same as in Fig. 1. For each thread, time is shifted so that the rank of the thread is 1 at time 0. The curve is $x_C(t)$ of (44) with the best fit $a = a^*$ and $b = b^*$ to the data. Horizontal and vertical axes are the hours and ranking, respectively.

In particular, we see a rather clear behavior close to the origin that the plotted points are on a curve tangential to y axis, indicating $x_C(t) = O(t^b)$ with $b < 1$ as in (45). Considering the simplicity of our model and formula, the fits seem good, suggesting a possibility of new application of fluid dynamics in the analysis of online rankings.

4.3 Amazon.co.jp book sales rankings.

We next turn to the ranking in the Amazon.co.jp online book sales. In this century of expanding online retail business, the economic impact of internet retails has attracted much attention, and there are studies using the sales rankings which appear on the webs of online booksellers such as Amazon.com [2, 3]. We will study Amazon.co.jp, a Japanese counterpart of Amazon.com, which seems to be not studied (and is easier to access for the authors). Amazon.com and Amazon.co.jp are similar in basic structures of web pages for individual books; on a web page for a book there are the title, price and related information such as shipping, brief description of the book, the sales ranking of the book, customer reviews and recommendations.

We should note that Amazon.co.jp, as well as Amazon.com, does not disclose exactly how it calculates rankings of books. In fact, there are observations [6] that Amazon.com defines the rankings for the top sales in a rather involved way. Therefore, it would be non-trivial and interesting if we could observe in the data behaviors similar to those of our simple model such as (44). See [5] for economic implications of the ranking numbers in Amazon.co.jp.

According to observation, Amazon.co.jp, as well as Amazon.com, updates their rankings once per hour, in contrast to the 2ch.net where the update procedure is instantaneous. This implies a limit of short time observational precision of 1 hour. On the other hand, for the long time observations, we have to consider a fact that the total number of books N is not constant. It is said that each year about 5×10^4 books are published in Japan, or about 5.7 books per hour. Certainly not all of the books are registered on Amazon.co.jp, so the increase of N per hour must be less than this value. Speed of ranking change decreases in the very tail side of the listing, and these practical changes in N will affect validity of applying (44) to the data in the very tail regime of the ranking. This gives a practical limit to long time analysis. Fortunately, at ranking as far down as 6.5×10^5, we still observe about 200 ranking change per hour, which makes an increase by 5.7 books negligible, so we expect a chance of applicability of our theory for long time data.

We will now summarize our results. The plotted 77 points in Fig. 3 show the result of our observation of a Japanese book rankings data, taken between the end of May, 2007 and mid August, 2007. As seen in the figure, the ranking number falls very rapidly near the top position (about 200 thousands in 5 days). The solid curve is a least square fit of these points to (44). Amazon.co.jp announces the total number of Japanese books in their list, which is a few times 10^6, but we suspect that this number includes a large number of books which are registered but never sell (so that we should discard in applying our theory). Therefore in addition to a and b in (44) we include N as a parameter to be fit from the data. Also, Amazon does not disclose the exact point of sales of each book, unlike 2ch.net where the exact jump time of a thread is recorded, so that the jump time to rank 1 of a book is also a parameter. The best fit for the parameter set (N, a, b) is:

$$(N^*, a^*, b^*) = (8.57 \times 10^5, 3.939 \times 10^{-4}, 0.6312), \quad (\sqrt{\chi^2/n_d} \simeq 1.4 \times 10^4).$$

Incidentally, we note in Fig. 3 a small jump at about 300 hours. We suspect this as a result of inventory control such as unregistering books out of print. Obviously, these controls need man-power, so that they appear only occasionally, making it a kind of unknown time dependent external source for our analysis.

All in all, we think it an impressive discovery that a simple formula as (44) could explain the data for more than 2 months. Our way of extracting basic sociological parameters such as the Pareto exponent b from the ranking data on the web has advantages over previous methods such as in [3], in that because the time development of the ranking of a book is a result of sales of as large number of books as $O(N)$, the book moves on the ranking queue in a deterministic way though each sale is a stochastic process. The fluctuations of sales (randomness about who buys what and when) are suppressed through a law-of-large-

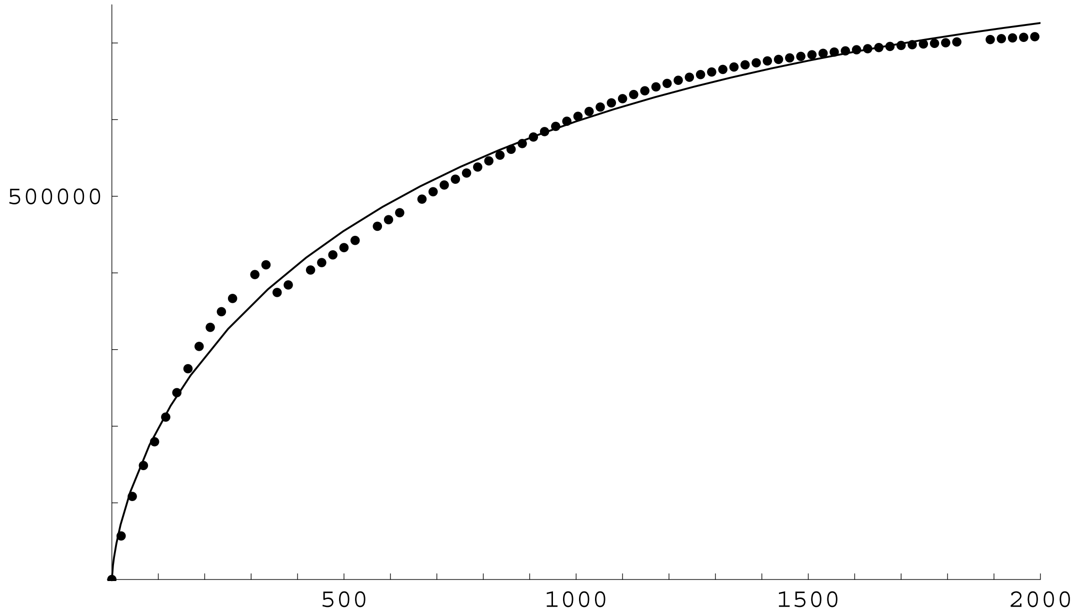

Fig 3: A long time sequence of data from Amazon.co.jp. The solid curve is a theoretical fit. Horizontal and vertical axes are the hours and ranking, respectively.

numbers type mechanism [4]. By looking at the time development of the ranking of a single book, we are in fact looking at the total sales of the books on the tail side of the book.

The theory of ‘long-tail economy’ says [1] that each product might sell only a little, but because of the overwhelming abundance in the species of the products the total sales will be of economic significance: It is not any specific single book but the total of books on the long-tail that matters. Our analysis on accumulated effect of products each with random and small sales, is particularly suitable in analyzing the new and rapidly expanding economic possibility of online retails, and moreover, is natural from the long-tail philosophy point of view.

Among the parameters to be fit in the Pareto distribution there is an exponent b which is of importance in the studies of economy. For example, in the case of distribution of incomes, which is usually quantitatively analyzed by the Pareto distribution, small b means that a few people of high incomes hold most of the wealth (the so called ‘20–80 law’ is a nickname for the Pareto distribution with $b = 1$), while for large b the society is more equal. In the case of ranking of online booksellers, large b means that there are many books (books in the ‘long-tail’ [1] regime), each of which does not sell much but the total sales of which is significant, further implying strong impacts of online retails to economy [3, 2], while small b favors dominance of traditional business model of ‘greatest hits’. Our studies on the 2ch.net bulletin board and the Amazon.co.jp online bookseller both consistently give the Pareto exponent $b \simeq 0.6$. Existing studies on online booksellers [3, 2] adopt the value of $b = 1.2$ and $b = 1.148$, respectively. These references also quote values from other studies, most of which satisfy $b > 1$. Note that [6] discovers, apparently based on extensive observations, that in the long-tail regime the sales are worse than in the head and intermediate regime, and gives the Pareto exponent $b = 0.4$ in the long-tail regime. Our method gives the total effect of intermediate and long-tails, so our value $b = 0.6$ could be more or less consistent with the observation of [6].

A Proof of equivalence of (10) and (4).

Here we prove that, if $\sup_i f_i < \infty$, (10) and (4) are equivalent, under the equations (1) (7) (2) (3) (5), with positive constants f_i, ρ_i, $i = 1, 2, \cdots$. (The extra condition on boundedness of f_i is of course irrelevant for the finite component cases.)

First assume (10). Then with (9) we have $v(0,t) = \sum_i f_i\,\rho_i$. On the other hand, integrating (1) by y from 0 to 1 and using (10) and (3) we have $v(0,t)\,u_i(0,t) = f_i\,\rho_i$. The two equations imply (4).

Next assume (4) and let

$$\overrightarrow{U} = (U_1, U_2, \cdots);\quad U_i(t) = \int_0^1 u_i(z,t)\,dz,\ \ i = 1, 2, \cdots. \tag{48}$$

Then (9) implies

$$v(0,t) = \sum_j f_j\,U_j(t). \tag{49}$$

Integrating (1) by y from 0 to 1, and using (3) (4) (49), we have

$$\frac{d\,\overrightarrow{U}}{dt}(t) = (A\,\overrightarrow{U})(t),\quad t \geqq 0, \tag{50}$$

with

$$(A\,\overrightarrow{U})_i(t) = \frac{f_i\rho_i}{\sum_j f_j\rho_j}\sum_j f_jU_j(t) - f_iU_i(t),\quad i = 1, 2, 3, \cdots,\ t \geqq 0. \tag{51}$$

The definition (5) implies $U_i(0) = \rho_i$, hence $U_i(t) = \rho_i$, $t \geqq 0$, is a solution, implying (10). To prove that this is a unique solution, let $\overrightarrow{U}$ be a solution of (50) with $\overrightarrow{U}(0) = \overrightarrow{\rho} := (\rho_1, \rho_2, \cdots)$. Then $\dfrac{d\,(\overrightarrow{U} - \overrightarrow{\rho})}{dt} = A\,(\overrightarrow{U} - \overrightarrow{\rho})$, hence

$$\sum_i |U_i(t) - \rho_i| \leqq \int_0^t \sum_{i,j} |A_{ij}||U_j(s) - \rho_j|\,ds.$$

Note that (51) implies $A_{ij} = \dfrac{f_i\rho_i f_j}{\sum_k f_k\rho_k} - \delta_{ij}f_i$, from which we have, noting the assumption $\sup_i f_i < \infty$,

$$\begin{aligned}\sum_i |U_i(t) - \rho_i| &\leqq \int_0^t 2\sum_i f_i|U_i(s) - \rho_i|\,ds \\ &\leqq 2\sup_i f_i \int_0^t \sum_i |U_i(s) - \rho_i|\,ds.\end{aligned}$$

Gronwall's inequality implies $\sum_i |U_i(t) - \rho_i| = 0$, $t \geqq 0$, hence $U_i(t) = \rho_i$, $i = 1, 2, \cdots$, $t \geqq 0$, is the unique solution to (50). This proves (10).

References

[1] C. Anderson, *The Long Tail: Why the Future of Business Is Selling Less of More,* Hyperion Books, 2006.

[2] E. Brynjolfsson, Y. Hu, M. D. Smith, *Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers,* Management Science **49-11** (2003) 1580–1596.

[3] J. Chevalier, A. Goolsbee, *Measuring prices and price competition online: Amazon.com and BarnesandNoble.com,* Quantitative Marketing and Economics, **1 (2)** (2003) 203–222.

[4] K. Hattori, T. Hattori, *Existence of an infinite particle limit of stochastic ranking processes,* Stochastic Processes and their Applications (2008), to appear.

[5] K. Hattori, T. Hattori, *Mathematical analysis of long tail economy using stochastic ranking processes,* preprint, `http://arxiv.org/abs/0804.0321`

[6] M. Rosenthal, `http://www.fonerbooks.com/surfing.htm`

M. Rosenthal, *Print-On-Demand Book Publishing: A New Approach to Printing and Marketing Books for Publishers and Self-Publishing Authors,* Foner Books, 2004.