# Sales ranks, Burgers-like equations, and least-recently-used caching

*Dedicated to Prof. K. R. Ito on his 60th birthday*

By

## Kumiko HATTORI* and Tetsuya HATTORI**

### Abstract

In this article, we summarize what we have found about the stochastic ranking process (move-to-front rule) and the sales ranks of online stores, as well as a system of Burgers-like partial differential equations and least-recently-used caching, in order to give an overview of relations among various topics in different fields.

## § 1.    Introduction.

In this article, we summarize what we have found about the stochastic ranking process (move-to-front rule) and the sales ranks of online stores, to give an overview of relations among various topics in different fields. For details including proofs of theorems and further references, we refer to [8, 9, 10, 11].

Amazon.co.jp, an online bookstore, gives each book it handles a number called 'Amazon sales rank,' which is renewed hourly on the store's websites. When one tracks the sales rank of a book that does not sell often, such as an academic book, one notices

*Department of Mathematics and Information Sciences, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan.
e-mail: `khattori@tmu.ac.jp`
**Laboratory of Mathematics, Faculty of Economics, Keio University, 4–1–1 Hiyoshi, Yokohama 223-8521, Japan.
URL: `http://web.econ.keio.ac.jp/staff/hattori/research.htm`
e-mail: `hattori@econ.keio.ac.jp`

wild behaviors, including occasional sudden jumps to much higher ranks (Fig. 2). We have proposed a stochastic model, which we called 'stochastic ranking process,' and have shown that it explains the observed peculiar behaviors well. It is a continuous-time version of a particle system following the move-to-front (MTF) rule, which has been known for half a century.

Combining our model with an assumption that the sales rates of books follow generalized Pareto distribution, we found that the way how a book's sales rank changes with time contains information on the whole business structure of Amazon's book division, not just the particular book's popularity. Specifically, a statistical fit of the observed data allows us to analyze the 'long tail' structure of Amazon.co.jp bookstore. Amazon has been known to be a pioneer of long-tail business, but their sales data have been beyond reach of researchers. Our result implies that the bookstore's dominant source of sales comes from a small number of bestsellers, rather than the collection of a huge number of unpopular books in the 'long tail' region.

Returning to the stochastic ranking process, we also proved that the infinite particle limit ('hydrodynamic limit') of the process exists and is deterministic. We further obtained an explicit formula for the limit. The limit can be characterized as the unique global solution to an initial value problem for a system of Burgers-like partial differential equations. A method of characteristic curves can be applied to solve the equations. In fact, it is these characteristic curves that we observe as the time development of sales ranks, which make possible to perform statistical fits of the web data to our formula. The method also mathematically explains the reason why the explicit formula is expressed in terms of inverse function of the characteristic curves. This viewpoint of partial differential equations and characteristic curves as its hydrodynamic limit seem to have been unnoticed for the process. The MTF rule has been studied as a model of LRU (least-recently-used) caching in computer science. Our result on 'hydrodynamic limit' gives a generalized formula of the cache miss probability.

## § 2.    Stochastic ranking process / Move-to-front rule.

We consider a system of $N$ particles lined in a queue. Each particle jumps to the top of the queue (rank 1) at random times. When a particle in rank $m$ jumps to rank 1, all the particles in ranks 1 through $m-1$ make a shift by one to ranks 2 through $m$ to fill the vacancy.

To formulate this model mathematically, let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $N$ be a natural number and $\mathcal{S}_N$ be the set of all permutations of $1, 2, \cdots, N$. We define the stochastic ranking process

$$X^{(N)}(t) = (X_1^{(N)}(t), \cdots, X_N^{(N)}(t)), \quad t \geqq 0,$$

as a Markov process with state space $\mathcal{S}_N$, as follows. Let $i = 1, 2, \cdots, N$ be labels identifying the particles and $X_i^{(N)}(t)$ be the rank (position in the queue) of particle $i$. For each $i = 1, \ldots, N$, there is an increasing sequence of random variables $\tau_{i,j}^{(N)}$, $j = 1, 2, \cdots$, such that $X_i^{(N)}(\tau_{i,j}^{(N)}) = 1$, $j = 1, 2, \cdots$; namely, the series of random times at which particle $i$ jumps to top. At a jump time $t = \tau_{i,j}^{(N)}$ of particle $i$, if particle $i'$ satisfies $X_{i'}^{(N)}(\tau_{i,j}^{(N)} - 0) < X_i^{(N)}(\tau_{i,j}^{(N)} - 0)$, then let $X_{i'}^{(N)}(\tau_{i,j}^{(N)}) = X_{i'}^{(N)}(\tau_{i,j}^{(N)} - 0) + 1$. so that $X^{(N)}(t) \in \mathcal{S}_N$ holds all the time. $X^{(N)}(t)$ changes only at jump times $\tau_{i,j}^{(N)}$. We assume that the initial state $X^{(N)}(0) = (X_1^{(N)}(0), \cdots, X_N^{(N)}(0))$ is given (deterministic), except in Section 6.

Fig. 1 is a sample configuration such that $N = 5$ and $\tau_{1,1}^{(5)} < \tau_{2,1}^{(5)} < \tau_{1,2}^{(5)} < \tau_{3,1}^{(5)}$. The first line in the figure illustrates $X_3^{(5)}(0) = 1$, $X_2^{(5)}(0) = 2$, $X_4^{(5)}(0) = 3$, $X_1^{(5)}(0) = 4$, $X_5^{(5)}(0) = 5$.
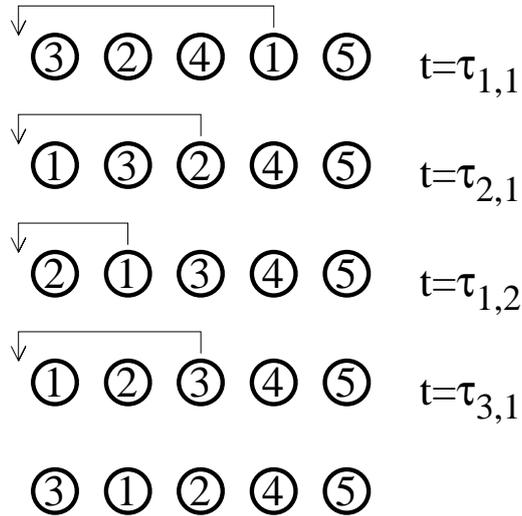


Fig 1. A sample configuration with $\tau_{1,1}^{(5)} < \tau_{2,1}^{(5)} < \tau_{1,2}^{(5)} < \tau_{3,1}^{(5)}$.

For notational simplicity, we put $\tau_{i,0}^{(N)} = 0$ for all $i = 1, 2, \cdots, N$, and we assume that $\{\tau_{i,j+1}^{(N)} - \tau_{i,j}^{(N)}, \ i = 1, 2, \cdots, N, \ j = 0, 1, 2, \cdots\}$ are independent and that for each $i = 1, 2, \cdots, N$, the jump intervals $\{\tau_{i,j+1}^{(N)} - \tau_{i,j}^{(N)}, \ j = 0, 1, 2, \cdots\}$ have an identical exponential distribution with parameter $w_i^{(N)} > 0$:

$$\mathrm{P}[\, \tau_{i,1}^{(N)} > t\,] = \exp(-w_i^{(N)}t).$$

Note that as in standard Poisson process, the jump times $\tau_{i,j}^{(N)}$, $j = 1, 2, \cdots$, $i = 1, 2, \cdots, N$, are distinct with probability 1. This completes the definition of the process $X^{(N)}$.

The stochastic ranking process defined above is a continuous-time version of the Markov chain known as the move-to-front (MTF) rule [19, 15, 12]  To be more precise, if we define $k$-th (discrete) jump time $\sigma^{(N)}(k)$ by

$$\{\sigma^{(N)}(k), \ k = 0, 1, 2, 3, \cdots\} = \{0\} \cup \{\tau_{i,j}^{(N)}, \ j = 1, 2, \cdots, \ i = 1, 2, \cdots, N\};$$
$$0 = \sigma^{(N)}(0) < \sigma^{(N)}(1) < \sigma^{(N)}(2) < \cdots,$$

we obtain a Markov chain $Z(k) = (X_1^{(N)}(\sigma^{(N)}(k)), \cdots, X_N^{(N)}(\sigma^{(N)}(k)))$ on $\mathcal{S}_N$. This Markov chain has been named in several ways such as move-to-front rules, self-organizing search and Tsetlin library [15, 12, 18, 19], and since the end of 20th century it has also often been referred to as least-recently-used (LRU) caching [14].

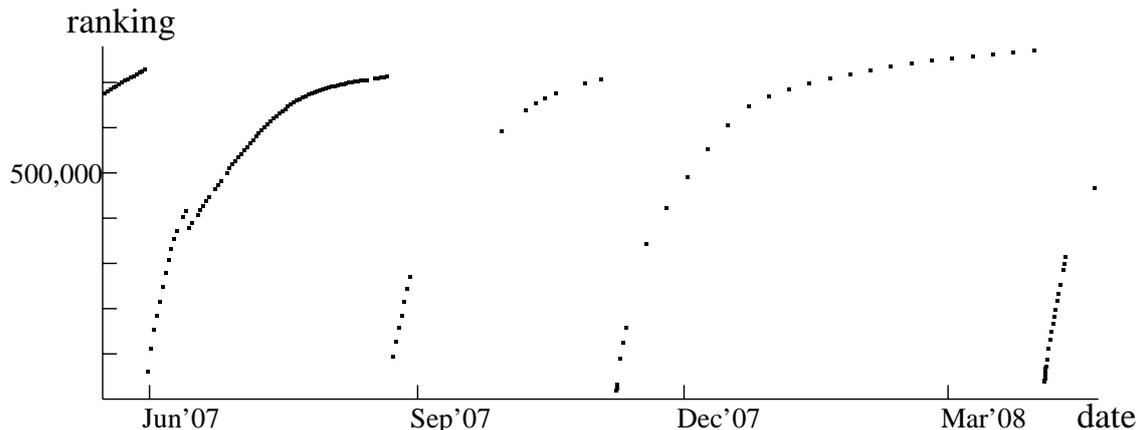## § 3.  Amazon.co.jp sales ranks.



Fig 2. A sample plot of Amazon.co.jp rankings for a book. The data were taken from May 15, 2007 to April 9, 2008. Note large discontinuous jumps to the top region (near the horizontal axis), which corresponds to the point of sales of the book. The data are taken manually; the density of the plot varies because it reflects how much the authors could devote their time on taking the data.

Amazon.co.jp is an online bookstore which is the Japanese counterpart of Amazon.com sales rank. On Amazon.co.jp's web pages, Amazon sales ranks are shown and renewed hourly. When one tracks the sales rank of any book that does not sell often (actually, most of the books on their catalog belong to this category), one notices peculiar behaviors. Most of the time, the rank stays around hundreds of thousands and keeps constantly falling (that is, the number representing the rank keeps increasing), but occasionally, it jumps up to as high as, say, ten thousand, which amounts to only a few percent of its usual rank and thus can be called 'the top area.' After such a large

jump, it starts falling again. Fig. 2 shows a plot of actual Amazon.co.jp ranks for a book taken over a year from May 15, 2007 to April 9, 2008. The original motivation for the stochastic ranking process defined in Section 2 was to explain the behaviors of these sales ranks. By actually ordering a copy at Amazon.co.jp website, it is easy to verify that a large jump occurs when someone orders the book at Amazon.co.jp. One or two hours after the order, a sudden jump in rank is actually observed. With this interpretation, we apply the stochastic ranking process to Amazon sales ranks.

Consider the scaled ranks

$$Y_i^{(N)}(t) := \frac{1}{N}(X_i^{(N)}(t) - 1) \in [0, 1),$$

and let

$$y_C^{(N)}(t) = \frac{1}{N}\sharp\{i \mid \tau_{i,1} \leqq t\} \in [0, 1).$$

$y_C^{(N)}(t)$ denotes a boundary of the head side and the tail side, where the head side consists of the particles which experienced jumps at least once by time $t$, and the particles in the tail side have not jumped at all up to time $t$. Denote the empirical distribution of jump rates by

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

where $\delta_a$ denotes a unit distribution concentrated at $a$.

**Proposition 3.1** ([8, Proposition 2]). *Assume that $\lambda^{(N)}$ converges weakly to some distribution $\lambda$ as $N \to \infty$. Then for each $t \geqq 0$,*

$$(3.1) \qquad y_C^{(N)}(t) \to y_C(t) := 1 - \int_0^\infty e^{-wt}\,\lambda(dw), \quad \text{in probability,}$$

*as $N \to \infty$. $y_C(t)$ is continuous and strictly increasing in $t$.* ◇

Proposition 3.1 is obtained as a straightforward consequence of the law of large numbers for independent random variables. Note that $y_C(t)$ is essentially the Laplace transform of $\lambda$.

Since the number $N$ of books in amazon.co.jp is very large ($N = O(10^6)$), we can apply this $N \to \infty$ result to sales ranks. In practical application to social and economical activities, a (generalized) Pareto distribution (also called a power law, a log–linear distribution, or a Zipf-like law for the discrete case) is often used. We assume this type of distribution for the probability distribution of book sales rate. Namely, we assume the probability measure $\lambda$ to be

$$(3.2) \qquad \lambda([w, \infty)) = \begin{cases} \left(\dfrac{a}{w}\right)^b, & w \geqq a, \\ 1, & w < a, \end{cases}$$

where, in terms of books in a bookstore, $w$ denotes the sales rate of a book title (average sales of $w$ copies per unit time), and $\lambda([w, \infty))$ is the ratio of the number of book titles with sales rate $w$ or more to the total number of books. The positive constant $a$ in (3.2) denotes the lowest positive sales rate of the books on the catalog of the bookstore. The other constant $b$ is also positive, where $-\dfrac{1}{b}$ is the so-called Pareto slope parameter. $N$ is the total number of book titles with positive sales rate. Note that the books that never sell should be ignored in applying the Pareto distribution (3.2).

Substituting (3.2) in (3.1) we have

$$(3.3) \qquad y_C(t) = 1 - ba^b \int_a^\infty e^{-wt} w^{-b-1} dw = 1 - b(at)^b \Gamma(-b, at),$$

where $\Gamma$ is the incomplete Gamma function defined by $\Gamma(z, p) = \displaystyle\int_p^\infty e^{-x} x^{z-1} dx$. Note in particular, that for $b < 1$ we have a concave time dependence for short time,

$$(3.4) \qquad y_C(t) = (at)^b \Gamma(1-b, 0) + o(t^b),$$

while for $b > 1$ we have linear short-time dependence. When $N$ is large enough,

$$(3.5) \qquad N\, y_C(t) = N\left(1 - e^{-at} + (at)^b \Gamma(1-b, at)\right)$$

approximates the sales rank of a book that started at rank 1 at time 0 and has not sold by time $t$.

We are interested in the value of $b$, which plays an important role in the analysis of business model. $b > 1$ implies that the long tail business is realized. By long tail, we refer to a business model where a huge number of low-sellers accumulate to bring in a considerable profit [1]. On the other hand, $b < 1$ implies that the business is best-seller based, that is, profit comes mostly from a small number of super hits.

Fig. 3 shows data from Amazon.co.jp ranking, and a statistical fit of the data to $Ny_c(t)$ in (3.5). The plots in Fig. 3 are the ranking data of a book in Amazon.co.jp on 21:00 JST each day between May 30, 2007 and August 16, 2007, one of the intervals between two adjacent jump to near rank 1 in Fig. 2. By performing statistical fits to the data we obtained $b = 0.81$ [10]. This result implies that Amazon.co.jp's business model is not long-tail based but best-seller based just like ordinary bookstores, in contrast to the idea in [1].

There are some previous studies where the value of $b$ for Amazon.com is estimated. Chevalier and Goolsbee had an estimate $b = 1.17$ [4], and Brynjolfsson, Hu and Smith obtained $b = 1.148$ [2]. Both implies long-tail business, as expected in [1]. In both studies, they used Amazon.com sales ranks to estimate $b$, but their estimate are based on the assumption that sales ranks are determined from the average sales. In our model,

ranking

500,000

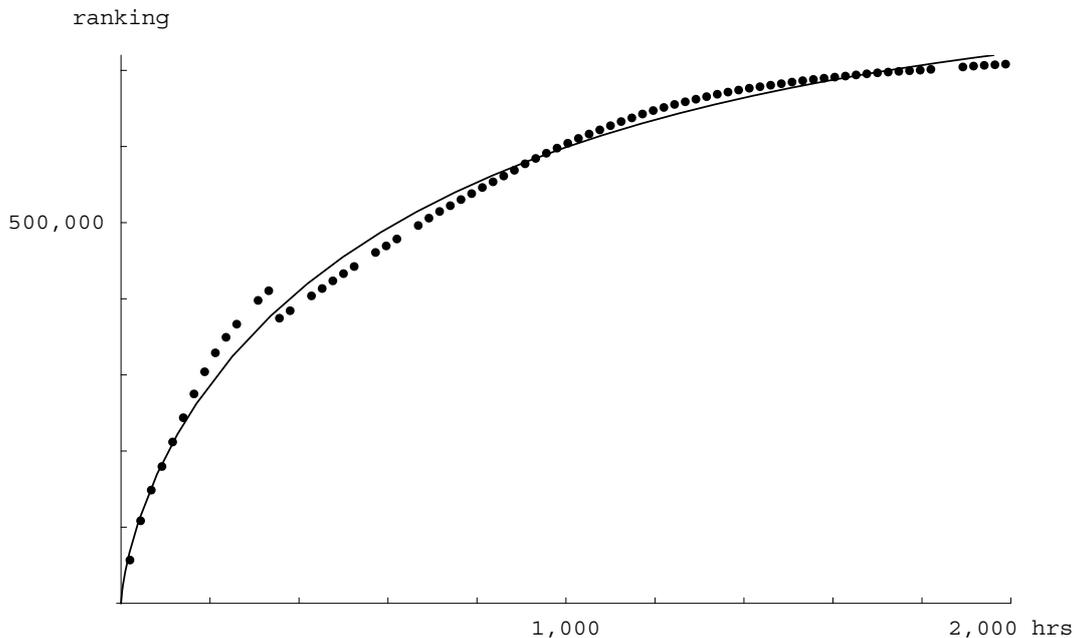1,000                              2,000 hrs

Fig 3. The horizontal axis represents time and the vertical axis sales ranks. The plots are the ranking data of a book in Amazon.co.jp on 21:00 JST each day between May 30, 2007 and August 16, 2007. The curve shows $Ny_C(t)$ obtained from a statistical fit to the data.

the rank depends on the time of the latest sale of a copy, rather than average sales. We only need to collect data for a single book, for the time development of a sales rank contains the information of the total sales of a large number of other books in the tail side. Note also that which book we choose for observation is theoretically irrelevant as long as it sells seldom enough to allow for a long-time observation.

We remark that values of $b$ such that $0 < b < 1$ have also been obtained in a study of document access in the MSNBC commercial news web sites [17] by direct measurements of access frequency.

It seems that the MTF ranking algorithm is not considered obvious for a model of sales ranks in real online stores; the authors have more than once received a response that sales ranks should be calculated in a more complicated algorithm. However as Fig. 2 shows, real sales ranks do behave wildly, and as Fig. 4 shows, the fit of the data to our theoretical curve $Ny_C(t)$ is quite good.
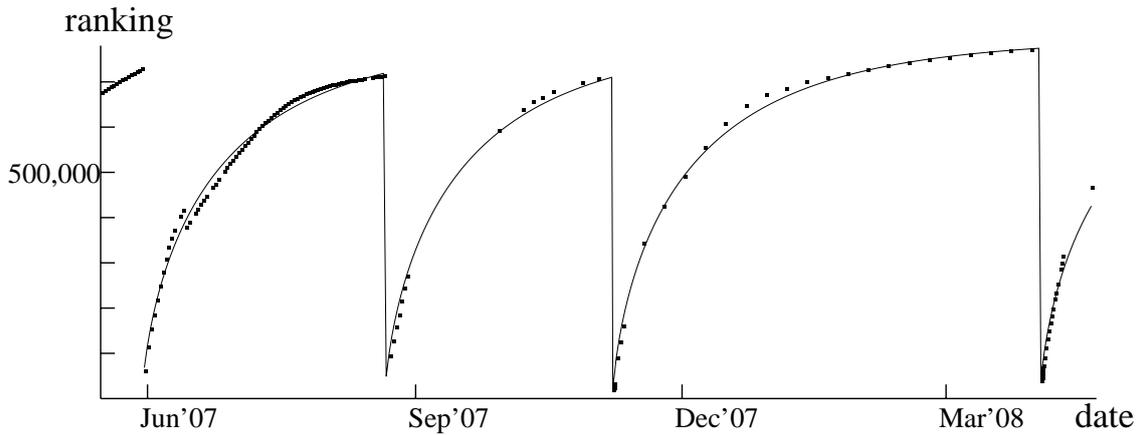
Fig 4. Plot of Amazon.co.jp rankings for a book from May 15, 2007 to April 9, 2008, as in Fig. 2, and the theoretical curve (3.5) with parameters as in Fig. 3.


## §4.  Infinite particle (hydrodynamic) limit.


Intuitively, one may guess that particles with large jump rates tend to stay near to the top of the ranks, while, if we have many particles, there will nearly always be some 'lucky' particles with small jump rates in the top area. The ratios of particles with different jump rates are random, but when $N$ is large enough we expect that the fluctuation is small. To formulate these intuition mathematically, let us define the joint empirical distribution (distribution-valued random variable) of jump rate and scaled rank at time $t$

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

We denote by $t_0(y)$ the inverse function of $y_C(t)$ given in Proposition 3.1, and also generalize the definition of $y_C$ to define

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

$y_C(y, t)$ is strictly increasing in $y$ and we denote the inverse function of $y_C(y, t)$ in $y$ by $\hat{y}(y, t)$.

**Theorem 4.1** ([8, Theorem 5]).  *Assume that $\lambda$ in Proposition 3.1 exists and satisfies $\int_0^\infty w\lambda(dw) < \infty$ and $\lambda(\{0\}) = 0$, and the initial joint distribution $\mu_0^{(N)}$ converges weakly to a distribution $\mu_0$ as $N \to \infty$. Then, for each $t > 0$, $\mu_t^{(N)}$ converges in*

*probability to a (non-random) distribution $\mu_t$ as $N \to \infty$. The limit $\mu_t$ is given by*

$$(4.1) \qquad U(dw, y, t) := \mu_t(dw, [y, 1)) = \begin{cases} \lambda(dw)\, e^{-wt_0(y)}, & y < y_C(t), \\ U(dw, \hat{y}(y, t), 0)\, e^{-wt}, & y > y_C(t), \end{cases}$$

*where $U(dw, y, 0) = \mu_0(dw, [y, 1))$.* $\qquad \diamond$

The assumption $\int_0^\infty w\lambda(dw) < \infty$ is required only in the proof of convergence at $y = 0$, and is irrelevant for $y > 0$.

## § 5.  Burgers-like equation.

Our proof of Theorem 4.1 is performed in such a way that first we infer the explicit form (4.1), and then directly prove that the difference between $\mu_t^{(N)}$ and the inferred $\mu_t$ converges to 0 (in the sense of distribution topology) as $N \to \infty$. In this proof, the explicit formula (4.1) is important. The formula can be mathematically characterized as the unique solution to the initial value problem of the following system of Burgers-like partial differential equations. Here we consider the case where jump rate takes at most countable different values $f_\alpha$, $\alpha = 1, 2, \cdots$ and $\lambda$ in Proposition 3.1 is given by

$$\lambda = \sum_\alpha \rho_\alpha \delta_{f_\alpha},$$

where the non-negative constant $\rho_\alpha$ denotes the the ratio of the particle of jump rate $f_\alpha$, and satisfies $\sum_\alpha \rho_\alpha = 1$.

**Theorem 5.1** ([9, Theorem 1]).  *$U_\alpha(y, t) := U(\{f_\alpha\}, y, t) = \mu_t(\{f_\alpha\}, [y, 1))$ is the unique time-global solution to the following initial value problem:*

$$(5.1) \qquad \begin{aligned} &\frac{\partial U_\alpha}{\partial t}(y, t) + \sum_\beta f_\beta\, U_\beta(y, t)\, \frac{\partial U_\alpha}{\partial y}(y, t) = -f_\alpha U_\alpha(y, t), \\ &(y, t) \in [0, 1) \times [0, \infty), \quad \alpha = 1, 2, \cdots, \end{aligned}$$

*with boundary condition*

$$U_\alpha(0, t) = \rho_\alpha, \ t \geqq 0, \ \alpha = 1, 2, \cdots,$$

*and for each $\alpha$, initial values $U_\alpha(y, 0) = U_\alpha(y)$, $0 \leqq y < 1$, are non-negative, differentiable, non-decreasing functions satisfying $\sum_\beta f_\beta U_\beta(0) < \infty$ and $\sum_\beta U_\beta(y) = 1 - y$.*

$\diamond$

Since (5.1) is a system of quasi-linear partial differential equations with common principal part, it can be solved in terms of the inverse functions of the characteristic curves $y_C$.

It seems that the existence and the explicit form of $\mu_t$ in Theorem 4.1 have not been noticed since the MTF model appeared in literature, and even for the quarter-century since the start of rather extensive studies in the application to computer science as a model of least-recently-used caching (Section 6). Perhaps this is because of the difficulty in finding the 'inverse function' $t_0(y)$ of the Laplace transform of $\lambda$ without the knowledge of its relation to the partial differential equation. Once one notices that the infinite particle limit is described by the partial differential equations as in the study of hydrodynamic limit, the inverse functions are obtained naturally by the method of characteristic curves in the case of (5.1). The essential key to the mathematical understanding of the $N \to \infty$ limit of the stochastic ranking process or MTF rules is to find the partial differential equations (5.1) characterizing the limit, and in this sense we may regard the limit Theorem 4.1 as a hydrodynamic limit.

We remark that the method of characteristic curves guarantees, in general, only the existence of local solutions. In the case of (5.1), thanks to the right hand side representing a loss of mass through evaporation, there is a shock-wave free condition

$$\sum_\beta \frac{\partial U_\beta}{\partial y}(y) \geqq -1,$$

which is satisfied by the initial data in Theorem 5.1, leading to existence of global solution.

## § 6.  Search cost.

There have been extensive studies on the search cost $C_N$ as an application of MTF rule, particularly in the study of computer sciences. (See, for example, [6, 13, 11], and references therein.) When a huge number of data are stored in a computer and accessed often, an efficient way is to copy a certain amount of frequently accessed data into a cache memory that allows quick access. Least-recently-used (LRU) caching is a simple algorithm of which data to keep in the cache memory. The algorithm is as follows: When a data not stored in the cache memory is accessed, it is reallocated in the cache memory, and in turn the data in the cache memory with the oldest access record is dropped from the cache. If one logically aligns the data in the order of latest access, one sees that this algorithm is equivalent to the MTF rule. In this application, one of the main interest is the position of an accessed data (the position of a particle just before its jump), which is the definition of $C_N$. If the ratio of the number of data in the

cache to that of all data is $y$, $\frac{1}{N}C_N \leqq y$ implies a cache success and $\frac{1}{N}C_N > y$ a cache miss (cache fault).

Denote the first particle to jump after time $t$ by $Q^{(N)}(t)$. Then the search cost is expressed as

$$\frac{1}{N}C_N = Y^{(N)}_{Q^{(N)}(t)}(t),$$

where $Y^{(N)}_\cdot(t)$ is defined in Section 2. Since particles jump independently of their positions, that is, events $\{Y^{(N)}_i(t) > x\}$ and $\{Q^{(N)}(t) = i\}$ are independent, the distribution of $\frac{1}{N}C_N$ can be calculated in terms of $\mu_t$ in Theorem 4.1, in the limit $N \to \infty$. We have

$$\lim_{N\to\infty} \mathrm{P}[\, \frac{1}{N}C_N(t) > x \,] = \frac{\iint_{(w,y)\in[0,\infty)\times(x,1)} w\,\mu_t(dw,dy)}{\int_0^\infty w\,\lambda(dw)}\,.$$

Although our motivation for Theorem 4.1 was a purely theoretical interest in the hydrodynamic limit and was independent of the trends in computer science, it turned out that the search cost can be expressed in a well-defined way in our notation, and Theorem 4.1 turned out to be applicable to the practical situations where the number of data $N$ is large [11].

Early works on MTF rule focused on the existence of the stationary distribution on $\mathcal{S}_N$ [19, 12] and the expected search cost at stationarity [15, 3]. When $N < \infty$, the general theory of Markov processes show that the stationary distribution on $\mathcal{S}_N$ is realized as $t \to \infty$. So far, we assumed that the initial configuration $(Y^{(N)}_1(0), \cdots, Y^{(N)}_N(0))$ is given (deterministic) in previous sections. In the following, let $(Y^{(N)}_1(0), \cdots, Y^{(N)}_N(0))$ be a $\mathcal{S}_N$-valued random variable with the stationary distribution. We denote the joint distribution of jump rate and position with the stationary initial configuration by $\mu^{(N)}_\infty$. Note that in Theorem 4.1, if $y < y_C(t)$, the limit distribution $\mu_t$ does not depend on the initial configuration or $t$. Thus it follows from Fubini's theorem and the dominated convergence theorem that $\mu_t$ is the limit distribution also of $\mu^{(N)}_\infty$ as $N \to \infty$, and

$$\lim_{N\to\infty} \mu^{(N)}_\infty(dw,dy) = \mu_t(dw,dy) = \frac{we^{-wt_0(y)}dy\lambda(dw)}{\int_0^\infty \tilde{w}e^{-\tilde{w}t_0(y)}\lambda(d\tilde{w})}\,.$$

This leads to formulas such as

$$\lim_{N\to\infty} \mathrm{P}[\, \frac{1}{N}C_N > x \,] = \frac{\int we^{-wt_0(x)}\lambda(dw)}{\int w\lambda(dw)}$$

in stationarity [11].

As noted in Section 5, it seems that a general formula expressed in terms of $t_0$ has not been obtained before, except asymptotic forms in the case of specific distributions.

For quantities which has a formula without inverse function $t_0$, we find many works. For example, the average search cost in the case of finite $N$ has been given by

$$\mathrm{E}_\infty[\, C_N \,] = \frac{1}{2} + \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{p_i^{(N)} p_j^{(N)}}{p_i^{(N)} + p_j^{(N)}}$$

in earlier works [15, 12], where we used the notation $\mathrm{E}_\infty$ to specify that we assume the stationary initial distribution. The limit of this formula as $N \to \infty$ is, as expected, equal to the search cost obtained using $\mu_\infty$ to give

$$\lim_{N\to\infty} \mathrm{E}_\infty[\, \frac{1}{N} C_N \,] = \frac{1}{\displaystyle\int_0^\infty w \, \lambda(dw)} \int_0^\infty \int_0^\infty \frac{w\tilde{w}}{w + \tilde{w}} \lambda(dw)\, \lambda(d\tilde{w}).$$

Among previous results for general distributions, a comparison between the average search cost $R_N$ for the optimal ordering (the case where the positions of data are fixed in the decreasing order of $w_i^{(N)}$) and that for the MTF case in stationarity is obtained as

$$\mathrm{E}[\, R_N \,] \leqq \mathrm{E}_\infty[\, C_N \,] \leqq 2\mathrm{E}[\, R_N \,] - 1.$$

In the limit, we have

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{E}[\, R_N \,] \leq \lim_{N\to\infty} \frac{1}{N}\mathrm{E}_\infty[\, C_N \,] \leq 2 \lim_{N\to\infty} \frac{1}{N}\mathrm{E}[\, R_N \,],$$

which can also be reproduced directly from our general result using

$$\frac{1}{2}\min\{x,y\} \leqq \frac{xy}{x+y} \leqq \min\{x,y\}, \ x,y \geqq 0.$$

We remark that [5] points out that using Hilbert's inequality, [7, §9.3], $K(x,y) = \dfrac{4xy}{(x+y)^3}$ with $p = q = 2$ and $g = f \geqq 0$,

$$\int_0^\infty \int_0^\infty \frac{4xy}{(x+y)^3} f(x)f(y)dxdy \leqq \frac{\pi}{2} \int_0^\infty f(x)^2 dx$$

a stronger (the best among those independent of $\lambda$) estimate from above

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{E}_\infty[\, C_N \,] \leqq \frac{\pi}{2} \lim_{N\to\infty} \frac{1}{N}\mathrm{E}[\, R_N \,]$$

is obtained.

## 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] P. J. Burville, J. F. C. Kingman, *On a model for storage and search,* J. Appl. Probab. **10** (1973) 697–701.

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

[5] F. R. K. Chung, D. J. Hajela, P. D. Seymour, *Self-organizing sequential search and Hilbert's inequalities,* J. Computer and System Sciences, **36** (1988) 148–157.

[6] J. A. Fill, *An exact formula for the move-to-front rule for self-organizing lists,* J. Theor. Probab. **9** (1996) 113–160.

[7] G. Hardy, J. E. Littlewood, G. Polya, *Inequalities,* Cambridge Univ. Press, 1988.

[8] K. Hattori, T. Hattori, *Existence of an infinite particle limit of stochastic ranking process,* Stochastic Processes and their Applications **119** (2009) 966–979.

[9] K. Hattori, T. Hattori, *Equation of motion for incompressible mixed fluid driven by evaporation and its application to online rankings,* Funkcialaj Ekvacioj (2009), **52** (2009) 301–319.

[10] K. Hattori, T. Hattori, *Mathematical analysis of long tail economy using stochastic ranking processes,* Preprint (2008), available at `http://web.econ.keio.ac.jp/staff/hattori/liamazn.htm`.

[11] K. Hattori, T. Hattori, *Hydrodynamic limit of move-to-front rules and search cost probabilities,* Preprint (2008), available at `http://web.econ.keio.ac.jp/staff/hattori/liamazn.htm`.

[12] W. J. Hendricks, *The stationary distribution of an interesting Markov chains,* J. Appl. Probab. **9** (1972) 231–233.

[13] P. R. Jelenković, *Asymptotic approximation of the move-to-front search cost distribution and least-recently used caching fault probabilities,* Ann. Appl. Probab. **9** (1999)430–464.

[14] P. R. Jelenković, A. Rdovanović, *Asymptotic insensitivity of least-recently-used caching to statistical dependency,* Proc. IEEE INFOCOM 2003, pp. 438–447, San Francisco, 2003.

[15] J. McCabe, *On serial files with relocatable records,* Oper. Res. **13** (1965) 609–618.

[16] M. Rosenthal, *What Amazon Sales Ranks Mean,* `http://www.fonerbooks.com/surfing.htm`, 2006.

[17] P. Qiu, *The content and access dynamics of a busy web site: Findings and implications,* Proc. of ACM SIGCOMM 2000, pp. 111–123, 2000.

[18] R. Rivest, *On self-organizing sequential search heuristics,* Comm. ACM **19** (1976) 63–67.

[19] M. L. Tsetlin, *Finite automata and models of simple forms of behaviour,* Russian Math. Surv. **18**(4) (1963) 1–27.