Where is my book? — Burgers equation in an online bookstore ranking

Tetsuya Hattori (Keio Univ.)
Kumiko Hattori (Tokyo Metropolitan Univ.)

We study a hydrodynamic limit approach to move-to-front rules, a stochastic process of particles aligned in a line, each of which jumps randomly to the top. We find that a scaling limit of the joint empirical distribution of jump rate and position satisfies a system of Burgers type partial differential equations, which in turn is globally solved using a standard method of characteristic curves.

Besides known application to LRU-caching in computer sciences, we find a new application in the ranking number of books at an online bookstore Amazon.co.jp. Statistical fit of data to our theory reveals that that the bookstore’s dominant source of sales is from a few bestsellers, rather than the majority of seldom sold books, contrary to the expectation of the ‘long tail’ business model.

1. Move-to-front rules. Let $N$ be a positive integer, and define a Markov process $X^{(N)}(t) = (X_1^{(N)}(t), \ldots, X_N^{(N)}(t))$, $t \geq 0$, with the state space a set of all permutations of $1, 2, \cdots, N$, as follows. $X^{(N)}(t)$ change values only at discrete random times (jump times) $\{\tau_{ij}^{(N)} | j = 1, 2, \cdots, i = 1, 2, \cdots, N\}$. At a jump time $t = \tau_{ij}$, we define $X_i^{(N)}(\tau_{ij}) = 1$ and, for $i' \neq i$,

$$X_{i'}^{(N)}(\tau_{ij}) = X_{i'}^{(N)}(\tau_{ij} - 0) + \begin{cases} 1, & \text{if } X_i^{(N)}(\tau_{ij} - 0) < X_{i'}^{(N)}(\tau_{ij} - 0), \\ 0, & \text{if } X_i^{(N)}(\tau_{ij} - 0) > X_{i'}^{(N)}(\tau_{ij} - 0). \end{cases}$$

We define the jump times to be independent in $i$, and for each $i$, the increments $\{\tau_{i,j+1}^{(N)} - \tau_{i,j}^{(N)} | j = 0, 1, 2, \cdots\} (\tau_{i,0}^{(N)} := 0)$ are i.i.d., whose distribution is the exponential distribution with parameter $w_i^{(N)} > 0$: $P[\tau_{i,1}^{(N)} > t] = \exp(-w_i^{(N)} t)$. Note that as in the Poisson process, with probability 1 the jump times are different for different $(i, j)$. In the following, we regard $X^{(N)}$ as an $N$ particle system aligned on a single line, with $X_i^{(N)}(t)$ denoting the position of the particle $i$ at time $t$.

2. Hydrodynamic limit. Consider a scaled position: $Y_i^{(N)}(t) := \frac{1}{N} (X_i^{(N)}(t) - 1) \in [0, 1)$. Note that $y_i^{(N)}(t) = \frac{1}{N} \mathbb{1}\{i | \tau_{i,1} \leq t\}$ is the rightmost point in scaled position of particles which have jumped to the top position up to time $t$. In the following we denote by $\delta_a$ the unit distribution concentrated on $a$, and assume

$$\lambda^{(N)} := \frac{1}{N} \sum_{i=1}^{N} \delta_{w_i^{(N)}} \to \lambda \text{ weakly as } N \to \infty, \text{ for a probability distribution } \lambda.$$ 

Proposition 1 ([1]). $y_i^{(N)}(t) \to y_C(t) := 1 - \int_0^\infty e^{-yt} \lambda(dw) \text{ } (N \to \infty, \text{ in prob.).}$

Consider a joint empirical distribution $\mu_t^{(N)} = \frac{1}{N} \sum_{i=1}^{N} \delta_{(w_i^{(N)}, Y_i^{(N)}(t))}$. 

Theorem 2 ([1]). Assume $\int_0^{\infty} w \lambda(dw) < \infty$ and $\lambda(\{0\}) = 0$, and assume that the initial distribution $\mu_0^{(N)}$ determined by the initial configuration $Y^{(N)}(0) = y^{(N)}$ converges weakly to a distribution $\mu_0$ as $N \to \infty$. Then for each $t > 0$, there exists a deterministic distribution $\mu_t$ such that $\mu_t^{(N)} \to \mu_t$ as $N \to \infty$. $\mu_t$ is given by

$$U(dw, y, t) := \mu_t(dw, [y, 1]) = \begin{cases} \lambda(dw) e^{-y \rho_0(y)}, & y < y_C(t), \\ U(dw, y, t, 0) e^{-yt}, & y > y_C(t), \end{cases}$$
where, \( t = t_0(y) \) is the inverse function of \( y = y_C(t) \), and \( \hat{y}(y, t) \) is the inverse function in \( y \) of \( y_C(y, t) = 1 - \int_1^y \int_0^\infty e^{-wt} \mu_0(dw, dz) \).

3. Burgers type equation. We found the explicit limit formula in Theorem 2 as a solution to a following system of PDEs. For simplicity of notation, consider the case where there are at most countable types of jump rates; \( \lambda = \sum_\alpha \rho_\alpha \delta_{f_\alpha} \), where \( f_\alpha \) and \( \rho_\alpha \) are positive constants, satisfying \( \sum_\alpha \rho_\alpha = 1 \).

**Proposition** 3([2]) \( U_\alpha(y, t) := U(\{f_\alpha\}, y, t) = \mu_t(\{f_\alpha\}, [y, 1)) \) is a unique classical time global solution to the following initial value problem:

\[
\frac{\partial U_\alpha}{\partial t}(y, t) + \sum_\beta f_\beta U_\beta(y, t) \frac{\partial U_\alpha}{\partial y}(y, t) = -f_\alpha U_\alpha(y, t), \ (y, t) \in [0, 1) \times [0, \infty), \ \alpha \in \mathbb{N},
\]

with boundary conditions \( U_\alpha(0, t) = \rho_\alpha, t \geq 0, \alpha \in \mathbb{N} \), and ‘nice’ initial data.

This system is solved by a standard method of characteristic curves, with explicit formula containing inverse functions such as \( t_0 \) of the characteristic curve \( y_C \). As in the theory of hydrodynamic limit, PDE explains a structure of the limit formula.

4. Amazon.co.jp. Move-to-front rule first appeared in the literature in 1963, and was extensively studied as a model of least-recently-used caching in the context of data theory in computer sciences, to which our results give some new formulas.

We moreover found a new application in the sales ranking of an online bookstore Amazon.co.jp. For each book which the bookstore sells, there is a web page on which is a number, updated every hour, called sales ranking of the book. The bookstore does not disclose the definition of the number, but from a continuous observation of the number for a less popular book (hmm, such as a book on mathematics...), we see that we may approximate its behavior by move-to-front rules (see the figure).

\[
\text{Note occasional large discontinuities, which correspond to the points that the book is purchased (as can be proved by ordering a book yourself). By performing a statistical fit of a continuous segment of ranking data to } y_C(t), \text{ choosing } \lambda \text{ as Pareto distribution, we deduced that the total sales of Amazon.co.jp is dominated by those of a few bestsellers (such as Harry Potter’s series) rather than a majority of unpopular books, contrary to the notion of ‘long tail’ business model, of which the Amazon bookstore has been considered as a pioneering company.}
\]

References.