Amazon ランキングと確率順位付け模型の流体力学極限

服部哲弥（慶應大・経済）

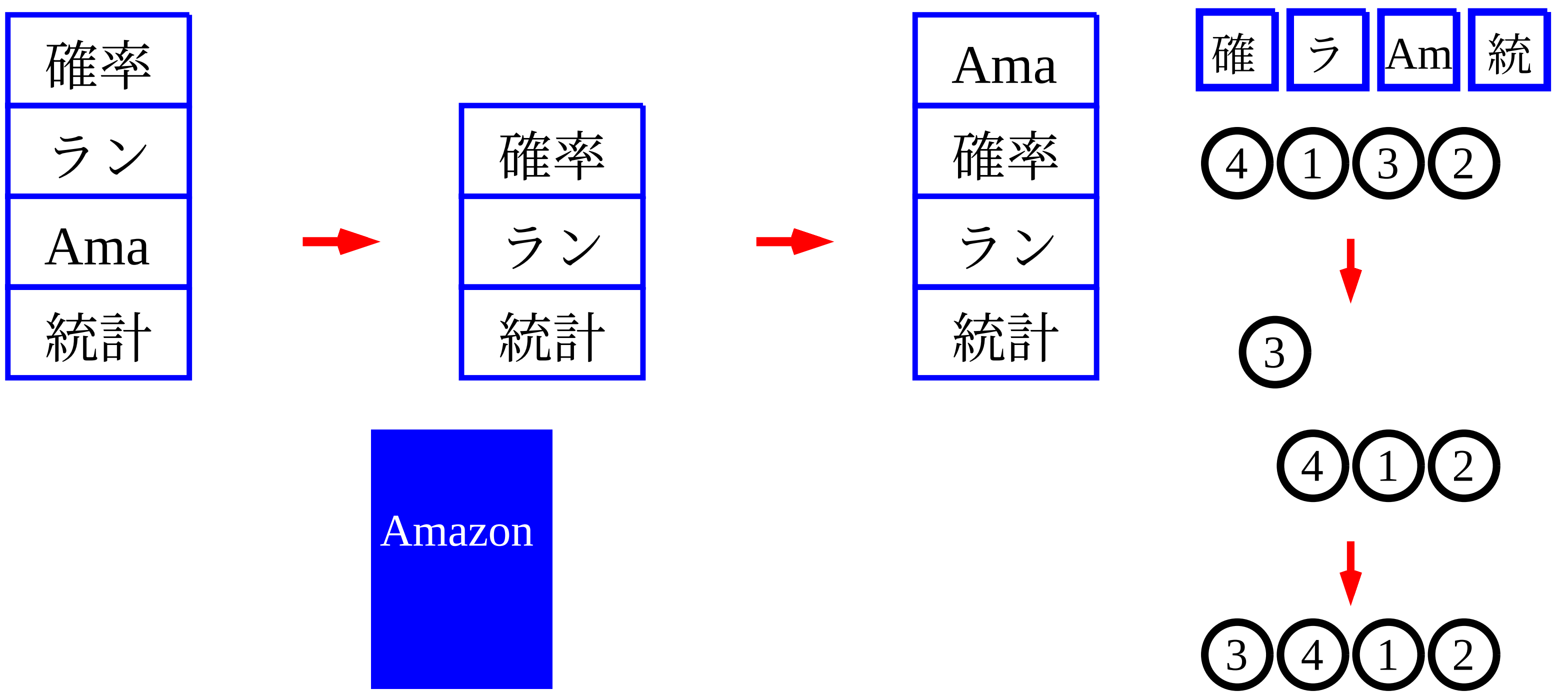
2019.09.17 日本数学会秋季総合分科会企画特別講演（金沢大学）

START ↱

０．目次

1．先頭に跳ぶ規則

「積ん読」：積み上げた本の塔（必要な本を都度抜き出し，終わったら塔の上に積む）動画１



Tsetlin（1963）**定常分布の算出（ノイズ下の有限状態オートマトンの例としてるが，多分下記１文のための例）**

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感を説明するかもしれない

ポワッソン過程による定式化

粒子数 N，粒子番号 $i = 1, \ldots, N$ （$N \to \infty$ を目指す），$T > 0$，時刻 $t \in [0, T]$

粒子系 $Y_i^{(N)}(t) \in \{\frac{i}{N} \mid i = 0, 1, \ldots, N-1\} \subset [0, 1)$，（順位は $N\,Y_i^{(N)}(t) + 1$）

初期値問題： $Y_i^{(N)}(0) = y_i^{(N)}$，$y_i^{(N)} \neq y_j^{(N)}$，$i \neq j$

先頭に跳ぶ規則：粒子 i が先頭 $(Y_i^{(N)}(\tau) = 0)$ に跳ぶとき追い越された粒子は順位を 1 ずつ下げる $(Y_j^{(N)}(\tau) = Y_j^{(N)}(\tau-) + \frac{1}{N})$

$\tilde{\nu}_i^{(N)}$: 独立なポワッソン過程（講演前半の簡単版），強度 w_i は i による

粒子 i が先頭に跳ぶ時刻＝$\tilde{\nu}_i^{(N)}$ の到着時刻

ー非減少非負整数値（点過程），$\nu(\tau) - \nu(\tau-) > 0$ なる τ を到着時刻

ー独立増分，$\mathrm{P}[\,\nu(t) - \nu(s) = k\,] = e^{-w(t-s)}\dfrac{w^k(t-s)^k}{k!}$，$k \in \mathbb{Z}_+$

確率順位付け模型

集合（事象）の定義関数 $\mathbf{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$

永幡幸生 (2013)

$$
\begin{aligned}
Y_i^{(N)}(t) = y_i^{(N)} &+ \frac{1}{N}\sum_{j=1}^{N}\int_0^t \mathbf{1}_{Y_j^{(N)}(s-)>Y_i^{(N)}(s-)}\,\tilde{\nu}_j^{(N)}(ds) \\
&- \int_0^t Y_i^{(N)}(s-)\tilde{\nu}_i^{(N)}(ds)
\end{aligned}
$$

右辺第３項：$\tilde{\nu}_i^{(N)}(\tau_{i,k}) - \tilde{\nu}_i^{(N)}(\tau_{i,k}-) = 1$　（到着時刻）$\Rightarrow$

$$
\begin{aligned}
Y_i^{(N)}(\tau_{i,k}) - Y_i^{(N)}(\tau_{i,k}-) &= -\int_{\tau_{i,k}-}^{\tau_{i,k}} Y_i^{(N)}(s-)\tilde{\nu}_i^{(N)}(ds) \\
&= -Y_i^{(N)}(\tau_{i,k}-)\left(\tilde{\nu}_i^{(N)}(\tau_{i,k}) - \tilde{\nu}_i^{(N)}(\tau_{i,k}-)\right) = -Y_i^{(N)}(\tau_{i,k}-)
\end{aligned}
$$

$\Rightarrow Y_i^{(N)}(\tau_{i,k}) = 0$（先頭に跳ぶ）

第２項：下位にいた粒子 j が先頭に跳ぶと i の順位が１下がる

$$
Y_i^{(N)}(\tau_{j,k}) - Y_i^{(N)}(\tau_{j,k}-) = \frac{1}{N}\int_{\tau_{j,k}-}^{\tau_{j,k}} \mathbf{1}_{Y_j^{(N)}(s-)>Y_i^{(N)}(s-)}\,\tilde{\nu}_j^{(N)}(ds) = \frac{1}{N}
$$

特性曲線

再掲：$Y_i^{(N)}(t) = y_i^{(N)} + \frac{1}{N}\sum_{j=1}^{N}\int_0^t \mathbf{1}_{Y_j^{(N)}(s-)>Y_i^{(N)}(s-)}\,\tilde{\nu}_j^{(N)}(ds)$

$$- \int_0^t Y_i^{(N)}(s-)\tilde{\nu}_i^{(N)}(ds)$$

第３項を除いた時間発展（先頭に跳ばない＝本の間に挟んだ「しおり」）

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

特に $y_0 = 0$ のとき $Y_C^{(N)}((0,t_0),t) = \frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_{\tilde{\nu}_j^{(N)}(t)>\tilde{\nu}_j^{(N)}(t_0)}$

独立確率変数列の平均　→　大数の強法則（大数の完全法則）

大数の完全法則

再掲（本の間に挟んだ「しおり」）：$Y_C^{(N)}((0,t_0),t) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\tilde{\nu}_j^{(N)}(t) > \tilde{\nu}_j^{(N)}(t_0)}$

再掲（ポワッソン）：$\mathsf{P}[\,\nu(t) > \nu(s)\,] = 1 - \mathsf{P}[\,\nu(t) - \nu(s) = 0\,] = 1 - e^{-w(t-s)}$

大数の完全法則　（分散有界なら異なる N の確率変数は総取り替えでかまわない）

$W = \mathbb{R}_+$ 上の分布 $\lambda^{(N)} := \frac{1}{N}\sum_{i=1}^{N} \delta_{w_i}$．$\lambda^{(N)} \to \lambda$ のとき，

$$y_C((0,t_0),t) := \lim_{N\to\infty} Y_C^{(N)}((0,t_0),t) = \lim_{N\to\infty} \mathsf{E}[\,Y_C^{(N)}((0,t_0),t)\,]$$

$$= \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \mathsf{P}[\,\tilde{\nu}_j^{(N)}(t) > \tilde{\nu}_j^{(N)}(t_0)\,]$$

$$= 1 - \int_W e^{-w(t-t_0)}\,\lambda(dw)$$

- ランダムな順位の時間発展が決定論的な時間発展で近似できる　動画２
- ポワッソン過程→指数関数の積分で（初等的に）書ける

2．Amazon ランキングの謎を解く

2010 年頃

Amazon.co.jp ランキング

amazon.co.jp

‘Internet retailers are extremely hesitant about releasing specific sales data’
（詳細を教えてくれないらしい）

ランキング（人気度，流行度）
これ以上ない簡単な数理モデル：

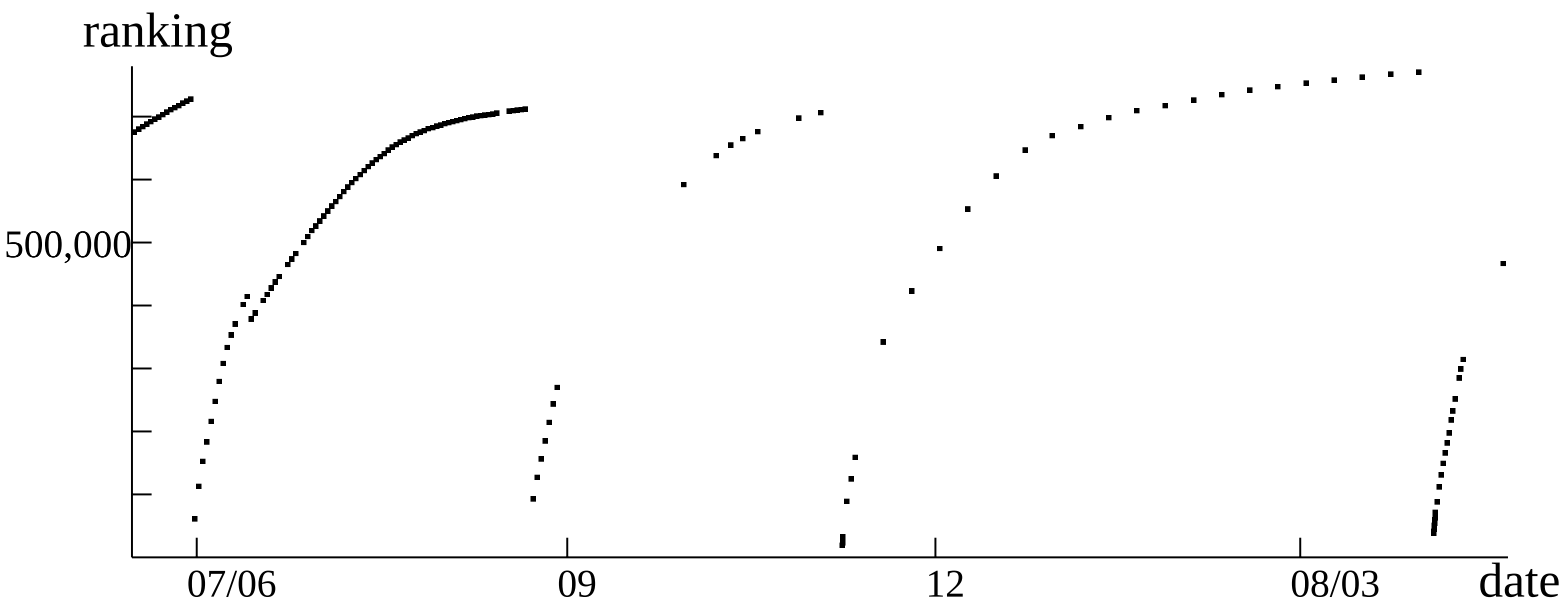
先頭に跳ぶ規則　＋
ポワッソン（指数）分布　＋
無限粒子極限

はどれくらい乱暴な近似か？

Amazon ランキングは先頭に跳ぶ！

Amazon ランキングと確率順位付け模型を（勝手に）対応させる：
粒子 i が先頭に跳ぶ時刻 $= \tilde{\nu}_i^{(N)}$ の到着時刻 ＝ 本 i がネット注文された時刻

注目する本がネット注文されるとランキングは１位に跳ね上がるか？



Amazon rankingへの確率順位付模型の当てはめ

再掲：粒子iが先頭に跳ぶ時刻＝$\tilde{\nu}_i^{(N)}$の到着時刻＝本iがネット注文された時刻

再掲：$y_C((0,t_0),t) = 1 - \int_W e^{-w(t-t_0)}\,\lambda(dw)$

時刻τに1位の本iが注文されぬまま時刻$t>\tau$になったときのランキング$X_i^{(N)}(t)$：

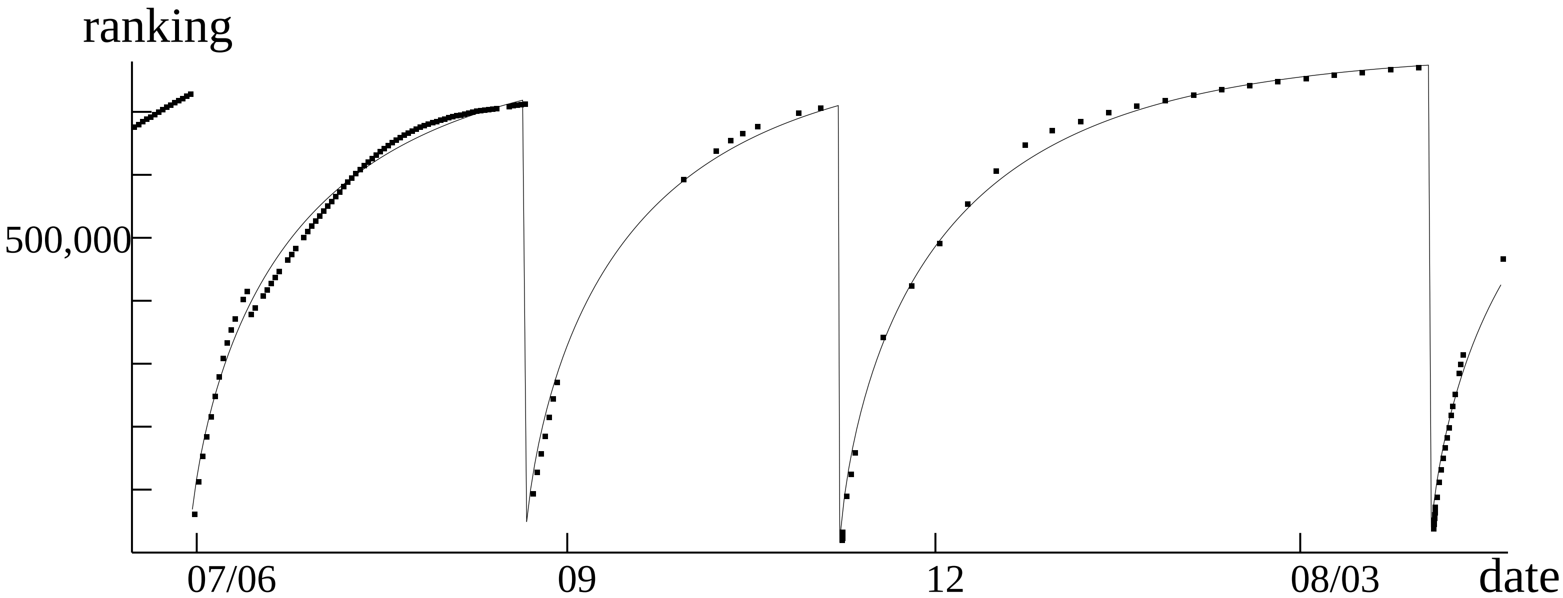
$$X_i^{(N)}(t) = N\,Y_i^{(N)}(t) + 1 \fallingdotseq N\,y_C((O,\tau),t) = N - N\int_0^\infty e^{-w\,(t-\tau)}\lambda(dw)$$

$\lambda = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_{w_i}$：注文頻度$w$の分布（売れる本売れない本の多寡）→部外秘情報

- **統計的当てはめ**：部外秘情報λを，公開情報だけで推定
- λをPareto分布で近似　→　ロングテールビジネスの検証

$\lambda([w,\infty)) = \left(\frac{a}{a\vee w}\right)^b$; a: 最低収入，b: 平等性の指数

アマゾンはロングテールに非ず



$$(N^*,\ a^*,\ b^*) = (8\times10^5,\ 6\times10^{-4},\ 0.81)$$

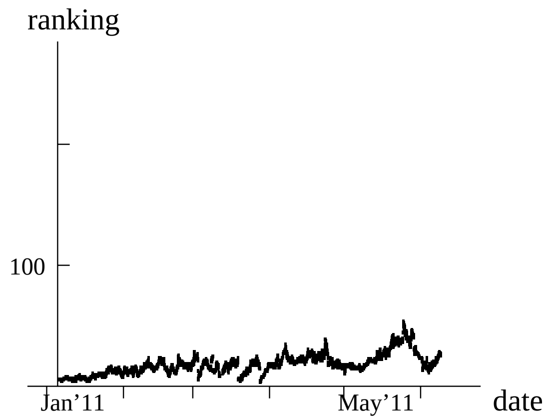
- $b<1$「不平等」上位 $o(N)$ 冊が売り上げの全てを占める

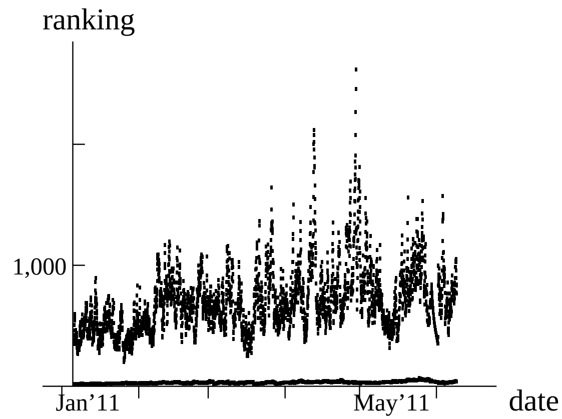
アマゾンはロングテールに非ず

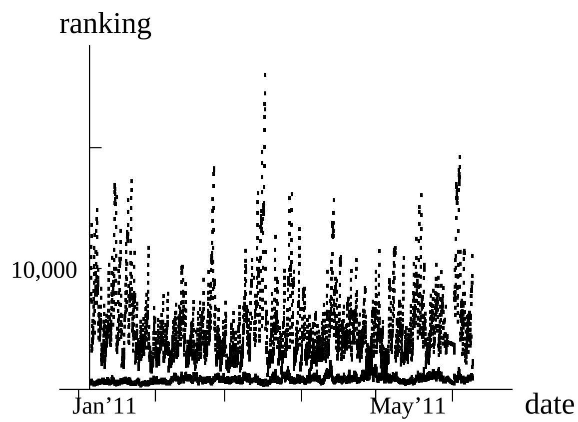
ロングテール型ではなく，ビッグヒット依存型のビジネスモデル

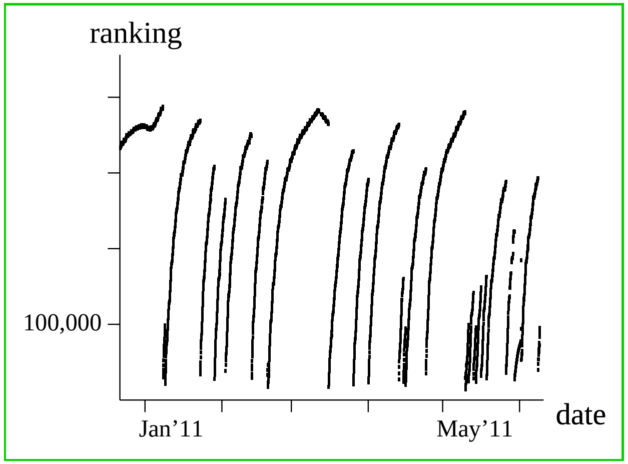
普通の本（ロングテール側）の順位変化

ビッグヒット vs. 普通の本









Amazonが勝手に決めているアルゴリズムが，単純なモデルと似た時間発展:奇跡的?
普遍性：殆どの本はめったに売れない．売れない本は最近売れた順になる．
ロングテール側（9割の本）の順位変化の普遍性　→　確率的順位付け

強度の時刻依存性

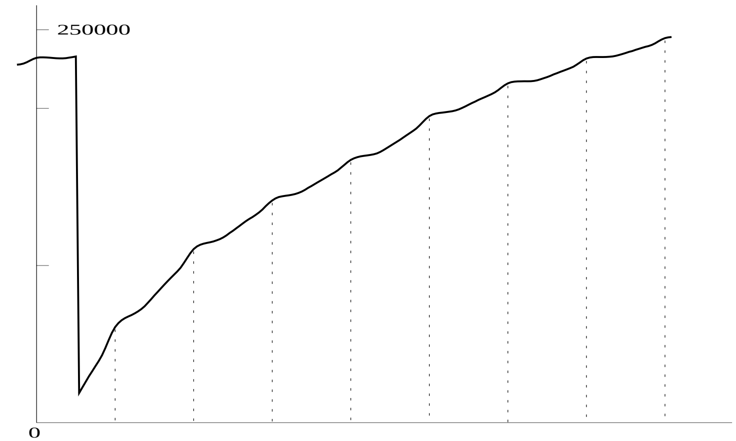
ここまで：w_i が定数のポワッソン過程 $\tilde{\nu}_i^{(N)}$

現実：昼夜差（毎時のランキングの自動取得）

時刻依存性がある強度を持つポワッソン過程 $\tilde{\nu}_i^{(N)}$

λ：関数の集合 W 上の分布とすれば〇K

$$w\,(t-t_0) \;\mapsto\; \int_{t_0}^{t} w(s)\,ds$$



λ: $W=\mathbb{R}_+$ 上の分布（定数強度）で当てはめが成立したのは？

共通の24時間周期 $w_i^{(N)}(t) = w_i^{(N)}\,A(t)$ を持つ場合に24時間毎にデータを取れば，$W=\mathbb{R}_+$ に帰着

位置強度結合経験分布

Tsetlin：定常分布

位置強度結合経験分布 $\mu_t^{(N)} = \frac{1}{N}\sum_{i=1}^{N}\delta_{(w_i, Y_i^{(N)}(t))}$ **の分布関数**

$$\varphi^{(N)}(dw,\gamma,t) = \mu_t^{(N)}(dw\times[Y_C^{(N)}(\gamma,t),1]),\ \ \gamma=(y_0,t_0)$$

$Y_C^{(N)}(\gamma,t)$ **と同様に粒子** i **が先頭に跳ぶという事象の和**

→　独立確率変数列の大数の完全法則

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

t_C **：増加関数** $y=y_C((0,0),t)$ **の逆関数**

$$t_C(y_C(\gamma,t)) = t - t_0 \text{ および } y = 1 - \int_W e^{-wt_C(y)}\lambda(dw)$$

定常分布 $\mu_t(dw\times[y,1)) = \mu_0(dw\times[y,1)) = e^{-wt_C(y)}\lambda(dw)$

あるいは $\mu_0(dw\times dy) = \dfrac{we^{-wt_C(y)}}{\int_0^\infty ve^{-vt_C(y)}\lambda(dv)}\lambda(dw)\,dy$

３．確率順位付け模型の流体力学極限

再掲．

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

強度 $w(t)$ が時刻依存性を持つ確率順位付け模型は方法結果とも定数強度の直接の拡張

$\tilde{\nu}_j^{(N)}$：（非一様）ポワッソン過程

分布関数 $\varphi^{(N)}(dw,\gamma,t) = \mu_t^{(N)}(dw \times [Y_C^{(N)}(\gamma,t),1])$
が独立確率変数列の平均（大数の完全法則）
独立増分→指数関数で書ける

問：位置依存性のある強度への一般化？
応用例：良い順位は宣伝効果があるか？
答：強度が位置依存性を持つ確率順位付け模型

強度が位置依存性を持つ確率順位付け模型

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

$\nu_i^{(N)}$: i について独立な，${\mathbb{R}_+}^2$ 上の単位強度のポワッソンランダム測度
$A \subset {\mathbb{R}_+}^2$ に対して $U(A) = \nu_i^{(N)}(A)$ は平均が面積のポワッソン分布に従う
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
w_i : $[0,1] \times [0,T] \to \mathbb{R}_+$: C^1 級

楠岡誠一郎 (2012)

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

- w を通した従属性

位置強度結合経験分布 $\mu_t^{(N)} = \dfrac{1}{N}\sum_{i=1}^{N}\delta_{(w_i,Y_i^{(N)}(t))}$ 　　流体力学極限：

「非ランダムな分布 μ_t があって，$N \to \infty$ で $\mu_0^{(N)} \to \mu_0$ を仮定すれば $[0,T]$ で $\mu_t^{(N)} \to \mu_t$」

流体力学極限

仮定：$\sup_{w\in W} \left\| \frac{\partial w}{\partial y} \right\|_{\mathsf{T}} < \infty$，強度分布 $\lambda^{(N)}(dw) = \mu_t^{(N)}(dw \times [0,1])$ の1次モーメント収束（w 非有界許すことはこだわりの1つ）

主定理. $N \to \infty$ で $\mu_0^{(N)} \to \mu_0$ ならば，$\exists \mu_t$; 確率1で t 一様に $\mu_t^{(N)} \to \mu_t$. さらに，$L \in \mathbb{N}$ と $y_i \in [0,1)$ $(1 \leqq i \leqq L)$ に対して $\nu_i^{(N)} = \nu_i$, $N \in \mathbb{N}$, と $\lim_{N\to\infty} y_i^{(N)} = y_i$ $(1 \leqq i \leqq L)$ とすると，確率1で $Y_i^{(N)}(t)$ は $N \to \infty$ で t 一様に収束 (propagation of chaos)

$$Y_i(t) = y_i + \int_{s\in(0,t]} \int_{(w,z)\in W\times[Y_i(s-).1]} w(z,s)\, \mu_s(dw \times dz)\, ds \\ - \int_{s\in(0,t]} \int_{\xi\in\mathbb{R}_+} Y_i(s-)\, 1_{\xi\in[0,w_i(Y_i(s-),s))}\, \nu_i(d\xi ds) \qquad \diamond$$

分布の集合の弱収束位相で時間について一様収束でポワッソンのサンプルについて概収束

妥当な仮定の下，期待する最善の結果　　極限は Poisson で書けない

強度が直前の到着時刻に依存する点過程

極限分布 $\varphi(dw,(y_0,t_0),t)=\mu_t(dw\times[y_C((y_0,t_0),t),1])$ **の良い記述→良い結果**
強度が直前の到着時刻に依存する点過程 $\tilde{\nu}$ **：到着時刻** $\tau_k=\inf\{t\geqq 0 \mid \tilde{\nu}(t)\geqq k\}$
$(\tau_0=0)$ **が，** $t\geqq\tau_{k-1}$ **で** $\mathrm{P}[\,t<\tau_k \mid \mathcal{F}_{\tau_{k-1}}\,]=\exp\left(-\int_{\tau_{k-1}}^{t} w(\tau_{k-1},u)\,du\right)$ **を満たす確率過程**

- **非独立増分**
- **ポワッソン過程との関係** $\mathrm{P}[\,\tilde{\nu}(t)=\tilde{\nu}(s)\,]=\sum_{k\geqq 0}\int_{0=:u_k<u_{k-1}<\cdots<u_1<u_0\leqq s}$

$$e^{-\sum_{i=0}^{k-1}\Omega(u_{i+1},u_i)-\Omega(u_0,t)}\left(\prod_{i=0}^{k-1}\omega(u_{i+1},u_i)du_i\right);\ \Omega(t_0,t)=\int_{t_0}^{t}\omega(t_0,u)du$$

ω **が第 1 変数によらない（位置不依存）とき**

$$\int_{0\leqq u_1\leqq u_2\leqq\cdots\leqq u_k\leqq s}\prod_{i=1}^{k} f(u_i)du_1\,du_2\ldots du_k=\frac{1}{k!}\left(\int_0^s f(v)dv\right)^k$$

によって指数関数 $\mathrm{P}[\,N(t)=N(s)\,]=e^{-\Omega(0,t)}e^{\Omega(0,s)}=e^{-\Omega(s,t)}$ **を再現**
展開してしまうと非有界な強度に対応できないが，公式や評価が使える

流れが定める強度に従う確率順位付け模型の極限

θ：初期値 $(y_0, 0)$・境界値 $(0, t_0)$ と時刻 t から位置 $[0, 1]$ への写像（で，y_C 類似）
Θ_T：流れ θ の集合
確率順位付け模型の強度 $w:\ [0,1]\times[0,T]\to\mathbb{R}_+$ と，流れ $\theta\in\Theta_T$ に対して
直前の到着時刻に依存する点過程 $\tilde{\nu}_{\theta,w,z}$ の強度 $w_{\theta,w,z}:\ [0,T]^2\to\mathbb{R}_+$ を

$$w_{\theta,w,z}(s,t)=\begin{cases} w(\theta((z,0),t),t), & \text{if}\quad s=0,\\ w(\theta((0,s),t),t), & \text{if}\quad s>0.\end{cases}$$

として，$\varphi_\theta(dw,(y_0,t_0),t):=\displaystyle\int_{z\in[y_0,1]}\mathrm{P}[\,\tilde{\nu}_{\theta,w,z}(t)=\tilde{\nu}_{\theta,w,z}(t_0)\,]\,\mu_0(dw\times dz)$

$\mathcal{G}(\theta)((y_0,t_0),t):=1-\varphi_\theta(W,(y_0,t_0),t)$ で $\mathcal{G}:\ \Theta_T\to\Theta_T$
定理． $\theta=\mathcal{G}(\theta)$ を満たす $\theta=y_C\in\Theta_T$ がただ一つ存在する ◇
極限分布 $\varphi(dw,(y_0,t_0),t)=\mu_t(dw\times[y_C((y_0,t_0),t),1])$ の答：$\varphi=\varphi_{y_C}$

特性関数で解ける偏微分方程式系の非局所項を持つ一般化

再掲： $y_C((y_0,t_0),t) = 1 - \varphi_{y_C}(W,(y_0,t_0),t)$

$$\begin{aligned}\varphi_{y_C}(dw,(y_0,t_0),t) &= \int_{z\in[y_0,1]} \mathsf{P}[\,\tilde{\nu}_{y_C,w,z}(t) = \tilde{\nu}_{y_C,w,z}(t_0)\,]\,\mu_0(dw\times dz)\\ &= \mu_t(dw\times[y_C((y_0,t_0),t),1])\end{aligned}$$

の言い換え．

流体成分 α **が有限種類のとき** $U_\alpha(y,t) = \mu_t(\{w_\alpha\}\times[y,1])$ **が満たす方程式：**

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

特性曲線： $\displaystyle \frac{d\,y_C}{dt}(t) = -\sum_\beta \int_{y_C(t)}^1 w_\beta(z,t)\,\frac{\partial U_\beta}{\partial z}(z,t)\,dz$

($U_\alpha(y_C(t),t) = \varphi_{y_C}(\{w_\alpha\},\gamma,t)$ **が** $\mathsf{P}[\,\tilde{\nu}_{y_C,w,z}(t) = \tilde{\nu}_{y_C,w,z}(t_0)\,]$ **で書ける)**

cf. w **が位置に依存しない場合は，空間** 1 **次元の特性曲線で解ける偏微分方程式系：**

$$\frac{\partial U_\alpha}{\partial t}(y,t) + \sum_\beta w_\beta(t)U_\beta(y,t)\,\frac{\partial U_\alpha}{\partial y}(y,t) = -w_\alpha(t)U_\alpha(y,t)$$

$$\frac{d\,y_C}{dt}(t) = \sum_\beta w_\beta(t)\,U_\beta(y_C(t)) \quad \rightarrow \quad U_\alpha(y_C(t),t) = U_\alpha(y_0,t_0)\,e^{-\int_{t_0}^t w_\alpha(s)ds}$$

主定理の証明のあらすじ

ざっくり言うと ...

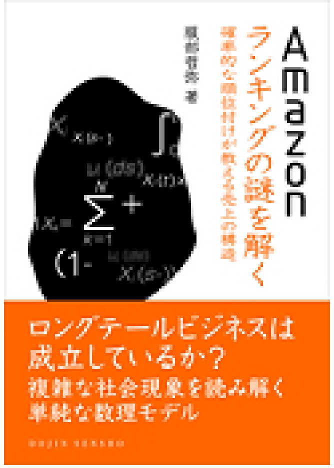
- **流れが定める強度に従う確率順位付け模型**
（中間模型：（放物型の極限を持つ系の）流体力学極限での局所平衡系の類推）
強度が直前の到着時刻に依存する（**非独立増分**な）点過程で粒子間では**独立**に先頭に跳ぶ → **大数の完全法則** $\varphi^{(N,\theta)} - \mathrm{E}[\,\varphi^{(N,\theta)}\,] \to 0$

- $\mathrm{E}[\,\varphi^{(N,\theta)}\,](dw,(y_0,t_0),t) \to \varphi_\theta(dw,(y_0,t_0),t)$ （←点過程の詳細性質）
（再掲：$\varphi_\theta(dw,(y_0,t_0),t) := \int_{z\in[y_0,1]} \mathrm{P}[\,\tilde{\nu}_{\theta,w,z}(t) = \tilde{\nu}_{\theta,w,z}(t_0)\,]\,\mu_0(dw\times dz)$）

- 確率順位付け模型模型と $\theta = y_C$ なる中間模型の差が $N \to \infty$ で消えることを全力証明 (Gronwall hierarchy)

詳しくは，**岩波「数学」最近号の論説**，および，ウェブ検索「Amazon ランキングの謎」

ウェブ検索キーワード：服部哲弥



T.Hattori, Tohoku Math. J. **71(3)** ('19)
T.Hattori, J.Math.Sci.U.Tokyo **25** ('18)
T.Hattori, Funkcialaj Ekvacioj **60** ('17)

服部哲弥,「Amazon**ランキングの謎を解く」化学同人**
服部哲弥,「確率変数の収束と大数の完全法則」共立出版

T.Hattori, S.Kusuoka, ALEA **9** ('12)
Y.Hariya, K.Hattori, T.Hattori, Y.Nagahata, Y.Takeshima, T.Kobayashi, Tohoku Math. J. **63** ('11)

K.Hattori, T.Hattori, Stoch.Proc.Appl. **119** ('09)