# ランキングプロセスに基づく
# オンラインランキングの分析

東北大学大学院 理学研究科 数学専攻
竹島　佑介

# 目 次

**第 1 章　序文** **2**

**第 2 章　ランキングプロセス** **5**
2.1　ランキングプロセスの定義 ： ： ：

# 第1章　序文

本修士論文では Amazon.co.jp などのオンラインブックストアの, 売上げランキングの時間発展をモデル化した**ランキングプロセス (stochastic ranking process)** に基づいて実際にオンラインランキングの分析を行った結果について述べる. ランキングプロセスとは, $N$ 個の粒子が点 1, 2, $\cdots$, $N$ に位置しているとして, 点 1 を先頭として先頭から何番目に位置しているかを**ランキング**と云うことにする. **ジャンプ時刻**に至るとその粒子はランキング 1 にジャンプして, ジャンプした粒子より前にいた粒子はそれぞれ 1 つずつランキングが下がる, ジャンプした粒子より後ろにいた粒子のランキングは変動しないという確率過程である. 先行研究として [4],[5] などではこの確率過程のことを Move-to-Front Rule という名で定義しており有限粒子のマルコフ連鎖としての考察をしている. ランキングプロセスという名は [1] で提唱されており, 無限粒子におけるランキングの時間発展は決定論的軌道になるということを考察している. ジャンプ経験済みの粒子の集合を head side, ジャンプ未経験の粒子の集合を tail side と云うことにすると, ジャンプの性質上 head side は粒子列の前方になり tail side は列の後方になる. この head side と tail side の境界のランキングを**境界値**と云うことにして, 時刻 $t$ での境界値を $x_C^{(N)}(t)$ と書くことにする. $x_C^{(N)}(t)$ は言い換えるとランキング 1 にジャンプしてから $t$ 分後のその粒子のランキングを表す. **無限粒子でのランキングプロセスを考察するにあたって粒子列を $[0,1)$ 区間に規格化する**. すなわち粒子は点 0, $\frac{1}{N}$, $\frac{2}{N}$, $\cdots$, $\frac{N-1}{N}$ に位置しているものとする. これに合わせて境界値を $y_C(t)$ と書くことにする. 粒子 $i$ のジャンプ時刻の分布を平均 $(w_i^{(N)})^{-1}$ の指数分布であるとすると, 定理 1.1 が従う. なお $w_i^{(N)}$ を**ジャンプ率**と云うことにする.

**定理 1.1** ([1, Proposition 2]). ジャンプ率の経験分布 $\quad^{(N)}$ は $N \to \infty$ のとき, $[0;\infty)$ 上のある分布 $\quad$ に弱収束しているとする. このとき

$$\lim_{N\to\infty} y_C^{(N)}(t) = 1 - \int_0^\infty e^{-wt}\ (dw) \quad \textit{in prob:}$$

が従う.

定理 1.1 の主張は**無限粒子ではランキングが 1 になってからのランキングの時間発展は決定論的になる**ということである. [2], [3] では, 定理 1.1 の主張を基にオンラインランキングでランキング 1 になってからのランキングの軌道を分析している. 本

論文では **2ch.net** に着目することにする．2ch.net とは電子掲示板の一つであり, それぞれの掲示板 (スレッド) を 板 と呼ばれるジャンルに分けている．各板にはスレッドの一覧表がありスレッドに書き込みがされるとその瞬間, そのスレッドは一覧表の先頭にジャンプする (ランキングが 1 になる) というまさにランキングプロセスと同じ振る舞いをする．[2] においても 2ch.net のスレッド一覧表でランキング 1 になってからのランキングの軌道を分析しているが, データ収集を手動で行っており 12 時間で分析していたため詳細な振る舞いまでは分かっていなかった．本論文ではプログラムを組むことでデータ収集を自動化して, 24 時間でランキングの時間発展の様子を分析することを第一の目標とした．その結果, ランキング 1 になってからのランキングの軌道は時間帯により異なることが分かった．すなわち実際のオンラインランキングではジャンプ率は一定ではなく時間帯により変動しているということが分かったため, この現象を解析するために理論を拡張してジャンプ率に時間依存性を入れた．この場合もジャンプ率一定のときと同じように無限粒子ではランキング 1 になってからのランキングの時間発展は決定論的軌道になる．

**定理 1.2** (定理 3.6 参照)**.** 粒子 $i$ のジャンプ時刻の法則を

$$\mathbf{P}[\ {}^{(N)}_{i} \leqq t] = 1 - \exp\left\{ - \int_0^t w_i^{(N)}(s) ds \right\}$$

であるとする．ただし時刻 $t$ における粒子 $i$ のジャンプ率を表す $w_i^{(N)}(t)$ は, 右連続かつ左極限を持つ関数であるとする．ジャンプ率の経験分布 $\Lambda^{(N)}$ は $N \to \infty$ のとき, ある分布 $\Lambda$ に弱収束しているとする．このとき

$$\lim_{N\to\infty} y_C^{(N)}(t) = 1 - \mathbf{E}^{\Lambda}\left[ \exp\left\{ - \int_0^t w(s) ds \right\} \right] \quad \textit{in prob:}$$

が従う．なお, $\mathbf{E}^{\Lambda}$ は $\Lambda$ に関する期待値を表す.

本論文では定理 1.2 を受けてジャンプ率は時間帯により変動するものとして, 2ch.net のスレッド一覧表ランキングの分析をやり直した．その結果実際のランキングの時間発展を分析するうえで統計的に改善がすることができた.

最後に本論文の構成を述べる．2 章ではランキングプロセスの定義を述べ, それに伴い現れる性質について述べる．3 章ではオンラインランキングのデータを収集し分析した結果を述べる．4 章ではジャンプ率と粒子位置の経験分布が無限粒子においてはルベーグ密度を持つ決定論的分布になることを述べる．付録 A ではオンラインランキングの分析を行うために作成したプログラムについて述べる．

## 謝辞

本修士論文を執筆するにあたり, 指導教官の服部哲弥先生には４年セミナーや院セミナー等で熱心な御指導を頂きました. また, 東北確率論セミナーを通じ竹田雅好先生, 針谷祐先生には貴重な御意見を頂きました. この場を借りて厚く御礼申し上げます. 先輩の田原喜宏さんには有益な助言を頂きました. 有り難う御座いました. 服部研究室の同輩の牛山辰哉君, 永沼伸顕君には, セミナーを含め議論に多数お付き合い頂きました. また, 私の友人である名古屋工業大学工学研究科所属の中村貴英君にはプログラミングに関する助言を頂きました. 深く感謝申し上げます.

最後に, 本人の希望を受け入れ勉強する環境を与えて頂いた父, 母に心からの感謝申し上げます.

# 第2章　ランキングプロセス

この章ではランキングプロセスの定義を述べそれに伴い現れる性質について述べる. 以下 $(\Omega; \mathcal{F}; \mathbf{P})$ を確率空間とする. また本論文を通して

- $\mathbb{R}_+ := [0; \infty)$,
- $C_b(I)$ : $I$ 上有界連続な関数全体の集合,
- $\#A$ : 集合 $A$ の濃度　特に $\#\emptyset = 0$,
- $_E(x)$ : 定義関数　すなわち $E \in \mathcal{F}$ に対し, $_E(x) = \begin{cases} 1; & x \in E \\ 0; & x \notin E \end{cases}$,
- $(x)$ : デルタ関数　すなわち $\int_E (x - x_0)dx = \begin{cases} 1; & x_0 \in E \\ 0; & x_0 \notin E \end{cases}$.

とする.

## 2.1　ランキングプロセスの定義

$N$ 個の粒子が点1, 点2, $\cdots$, 点$N$ に並んでいるとする. 点1にある粒子を先頭として, 先頭から何番目にあるかを**ランキング**と云うことにする. 粒子 $i$ $(i = 1; 2; \cdots ; N)$ の初期ランキングを $x_{i,0}^{(N)}$ で表すことにする. ただし $\{x_{i,0}^{(N)} \mid i = 1; 2; \cdots ; N\}$ は $\{1; 2; \cdots ; N\}$ の順列であるとする. $j = 0; 1; 2; \cdots$ に対し, 粒子 $i$ の第 $j$ **ジャンプ時刻**を $_{i,j}^{(N)} : \Omega \to \mathbb{R}_+$ で定義する.

**定義 2.1.** $(\Omega; \mathcal{F}; \mathbf{P})$ 上の確率過程 $\{X_i^{(N)}(t) \mid t \geqq 0;\ i = 1; 2; \cdots ; N\}$ を以下の (2.1.1)-(2.1.3) で定義して**ランキングプロセス**と云う.

(1) $0 \leqq t < {}_{i,1}^{(N)}$ のとき

$$X_i^{(N)}(t) = x_{i,0}^{(N)} + \#\{i' \in \{1; 2; \cdots ; N\} \mid x_{i',0}^{(N)} > x_{i,0}^{(N)};\ {}_{i',1}^{(N)} \leqq t\}: \tag{2.1.1}$$

(2) $t = {}_{i,j}^{(N)}$, $j = 1; 2; \cdots$ のとき

$$X_i^{(N)}(t) = 1: \tag{2.1.2}$$

(3) $\tau_{i,j}^{(N)} < t < \tau_{i,j+1}^{(N)}$, $j = 1;2;\cdots$ のとき

$$X_i^{(N)}(t) = \#\{i' \in \{1;2;\cdots;N\} \mid \exists j' \in \mathbb{N};\ \tau_{i,j}^{(N)} \leqq \tau_{i',j'}^{(N)} \leqq t\}: \tag{2.1.3}$$

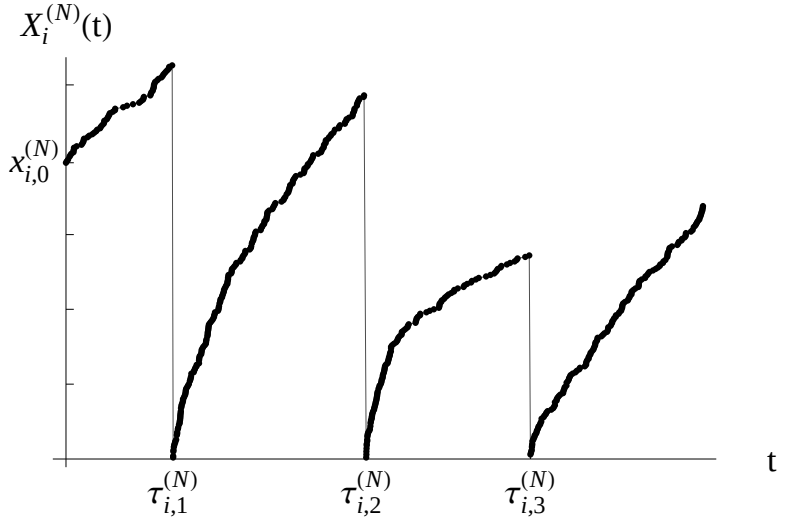


図 2.1: $X_i^{(N)}(t)$ のグラフ

**注 2.2.** 定義 2.1 を言い換えると粒子 $i$ のランキングは時刻 $t = 0$ で $x_{i,0}^{(N)}$ であり, ジャンプ時刻に至った瞬間その粒子のランキングは 1 になる. 自粒子のジャンプ時刻間は, 自粒子よりもランキングが低い粒子のジャンプ時刻に至るごとに 1 つずつランキングが下がっていく. この際他粒子を追い越して後ろに下がることは無い. また $\{x_{i,0}^{(N)} \mid i = 1;2;\cdots;N\}$ は $\{1;2;\cdots;N\}$ の順列であることから, 各時刻 $t$ で $\{X_i^{(N)}(t) \mid i = 1;2;\cdots;N\}$ は $\{1;2;\cdots;N\}$ の順列になる.

ジャンプ時刻の列 $\{\tau_{i,j}^{(N)} \mid j = 0;1;2;\cdots\}$ について次を仮定する.

**仮定 2.3.** $i = 1;2;\cdots;N$ に対し $w_i^{(N)}$ は正定数とする.

(1) $\tau_{i,0}^{(N)} = 0$,

(2) $\{\tau_{i,j}^{(N)} \mid j = 0;1;2;\cdots\}$, $i = 1;2;\cdots;N$ は独立,

(3) $\tau_i^{(N)} := \tau_{i,1}^{(N)}$ の法則は平均が $(w_i^{(N)})^{-1}$ の指数分布　すなわち $t \geqq 0$ に対して

$$\mathbf{P}[\tau_i^{(N)} \leqq t] = 1 - e^{-w_i^{(N)}t}; \tag{2.1.4}$$

を満たすものとする.

**注 2.4.** 各 $i$ に対し, $\{ {}^{(N)}_{i,j} \mid j = 0; 1; 2; \cdots \}$ は増大列で, (2.1.4) より確率 1 で狭義増大列である. また $(i; j) \neq (i'; j')$ に対し, 確率 1 で ${}^{(N)}_{i,j} \neq {}^{(N)}_{i',j'}$ である.

**命題 2.5** ([1, Proposition 1])**.** 各 $i = 1; 2; \cdots ; N$ に対し次が従う.

(1) $j = 1; 2; \cdots$ に対し, $X_i^{(N)}( {}^{(N)}_{i,j}) = 1$.

(2) $i' \neq i$, $j' = 1; 2; \cdots$ に対し, $X_i^{(N)}( {}^{(N)}_{i',j'} - 0) < X_{i'}^{(N)}( {}^{(N)}_{i',j'} - 0)$ ならば

$$X_i^{(N)}( {}^{(N)}_{i',j'}) = X_i^{(N)}( {}^{(N)}_{i',j'} - 0) + 1:$$

(3) $i' \neq i$, $j' = 1; 2; \cdots$ に対し, $X_i^{(N)}( {}^{(N)}_{i',j'} - 0) > X_{i'}^{(N)}( {}^{(N)}_{i',j'} - 0)$ ならば

$$X_i^{(N)}( {}^{(N)}_{i',j'}) = X_i^{(N)}( {}^{(N)}_{i',j'} - 0):$$

**例 2.6.** $N = 7$ とし, 初期時刻に図 2.2 のように粒子が並んでいるとし, $0 < {}^{(7)}_{4} < {}^{(7)}_{3} < {}^{(7)}_{6} < {}^{(7)}_{2}$ であるとする. 時刻 ${}^{(7)}_{4}$ になった瞬間粒子の並びは図 2.3 になる. さらに, ${}^{(7)}_{2}$ になった瞬間粒子の並びは図 2.4 になる.

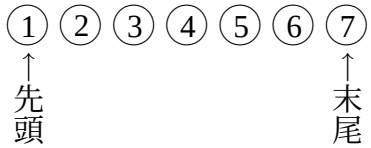


図 2.2: 初期時刻における粒子列

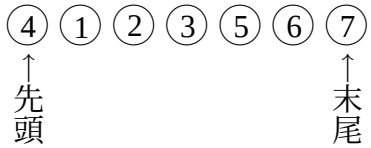


図 2.3: 時刻 ${}^{(7)}_{4}$ における粒子列

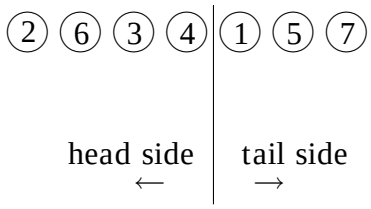


図 2.4: 時刻 ${}^{(7)}_{2}$ における粒子列

便宜上区間 [0,1) にスケールしたランキングプロセスを考えることもあるためここで定義しておく.

**定義 2.7.** ランキングプロセス $\{X_i^{(N)}(t) \mid t \geqq 0;\ i = 1;2;\cdots;N\}$ に対し,

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

で定義される $\{Y_i^{(N)}(t) \mid t \geqq 0;\ i = 1;2;\cdots;N\}$ を**スケールしたランキングプロセス**と云う. 特に混乱がない場合は $Y_i^{(N)}(t)$ も単にランキングプロセスと云うことにする. また, これに合わせ $y_{i,0}^{(N)} := \frac{1}{N}(x_{i,0}^{(N)} - 1)$ と定義する.

## 2.2 境界値

各時刻 $t$ において, $t < {}_i^{(N)}$ ならば粒子 $i$ はジャンプ未経験であり, $t \geqq {}_i^{(N)}$ ならば粒子 $i$ はジャンプ経験済みである. ジャンプ未経験の粒子のランキングは他の粒子のジャンプ時刻をむかえるごとに下がっていく. ゆえに粒子列の前方にはジャンプ経験済みの粒子が集まり, 後方にジャンプ未経験の粒子が集まる.

**定義 2.8.** 時刻 $t$ における tail side での先頭粒子のランキングを $x_C^{(N)}(t)$ と書くことにし, **境界値**と云う. ただし全ての粒子が head side に属すなら $x_C^{(N)}(t) = N + 1$ とする. すなわち $x_C^{(N)}(t) : \Omega \to \{1;2;\cdots;N+1\}$ を

$$x_C^{(N)}(t) := \sum_{i=1}^{N} {}_{\tau_i^{(N)} \leqq t} + 1$$

で定義する. また区間 [0,1) にスケールしたものの境界値を

$$y_C^{(N)}(t) := \frac{1}{N}(x_C^{(N)}(t) - 1) = \frac{1}{N}\sum_{i=1}^{N} {}_{\tau_i^{(N)} \leqq t} \tag{2.2.1}$$

で定義する.

**注 2.9.** 定義 2.8 を言い換えると初期時刻にランキング 1 の粒子が時刻 $t$ において tail side にいるのなら, この粒子の時刻 $t$ におけるランキングは $x_C^{(N)}(t)$ である.

仮定 2.3(3) より $w_i^{(N)}$ は粒子 $i$ の**ジャンプ率**を表わす. ジャンプ率の経験分布を

$${}^{(N)}(dw) := \frac{1}{N}\sum_{i=1}^{N} (w - w_i^{(N)})dw$$

と置く. $N \to \infty$ のとき, $y_C^{(N)}(t)$ は 決定論的極限に収束する.

**定理 2.10** (定理1.1再掲). $\mu^{(N)}(dw)$ は $N \to \infty$ のとき, $\mathbb{R}_+$ 上のある分布 $\mu(dw)$ に弱収束しているとする. すなわち任意の $g \in C_b(\mathbb{R}_+)$ に対し,

$$\lim_{N\to\infty} \int_0^\infty g(w)\ \mu^{(N)}(dw) = \int_0^\infty g(w)\ \mu(dw) \tag{2.2.2}$$

であるとする. このとき

$$\lim_{N\to\infty} y_C^{(N)}(t) = y_C(t) \quad \textit{in prob:}$$

が従う. ただし,

$$y_C(t) = 1 - \int_0^\infty e^{-wt}\ \mu(dw) \tag{2.2.3}$$

である.

この命題の証明及び, 以下で大数の弱法則を幾度か用いるためここで述べておく.

**補題 2.11** (大数の弱法則). $\{X_n\}$ を確率空間 $(\Omega; \mathcal{F}; \mathbf{P})$ 上の独立な確率変数列とする. 任意の $n \in \mathbb{N}$ に対し, $\mathbf{E}[|X_n|^2] < \infty$ かつ $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{N} \mathbf{E}[X_n] = a$ であるとする. このとき

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = a \quad \textit{in prob:}$$

が従う.

**証明**. 任意の $\varepsilon > 0$ に対し,

$$\begin{aligned}
\mathbf{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| > \varepsilon\right] &= \mathbf{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbf{E}[X_i]) + \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i] - a\right| > \varepsilon\right] \\
&\leqq \mathbf{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbf{E}[X_i])\right| > \frac{\varepsilon}{2}\right] \\
&\quad + \mathbf{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i] - a\right| > \frac{\varepsilon}{2}\right];
\end{aligned}$$

チェビシェフの不等式を用いることで,

$$\begin{aligned}
\mathbf{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbf{E}[X_i])\right| > \frac{\varepsilon}{2}\right] &\leqq \left(\frac{2}{n\varepsilon}\right)^2 \mathbf{E}\left[\left|\sum_{i=1}^{n}(X_i - \mathbf{E}[X_i])\right|^2\right] \\
&= \left(\frac{2}{n\varepsilon}\right)^2 \mathrm{Var}\left[\sum_{i=1}^{n}(X_i - \mathbf{E}[X_i])\right]
\end{aligned}$$

$$= \left(\frac{2}{n''}\right)^2 \sum_{i=1}^{n} \mathrm{Var}[X_i]$$
$$\to 0 \quad \text{as } n \to \infty:$$

より補題 2.11 を得る. □

**定理 2.10 の証明.** 仮定 2.3 (2) より $\{ \ _{\tau_i^{(N)} \leqq t} \mid i = 1; 2; \cdots ; N \}$ は独立である. (2.1.4) より

$$\begin{aligned} \mathbf{E}[\ ^{2}_{\tau_i^{(N)} \leqq t}] &= \mathbf{P}[\ ^{(N)}_{i} \leqq t] \\ &= 1 - e^{-w_i^{(N)} t} \\ &< 1 \end{aligned}$$

が従う. また (2.2.2) より

$$\begin{aligned} \lim_{N\to\infty} \mathbf{E}[y_C^{(N)}(t)] &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \int_0^{\infty} (1 - e^{-wt})\ (w - w_i^{(N)}) dw \\ &= \lim_{N\to\infty} \int_0^{\infty} (1 - e^{-wt})\ \ ^{(N)}(dw) \\ &= y_C(t) \end{aligned}$$

が従うので, 補題 2.11 より定理 2.10 を得る. □

# 第3章　オンラインランキングの分析

この章では2章で定義したランキングプロセスを基にオンラインランキングの分析をした結果について述べる．まずランキングプロセスの例として

- Amazon.co.jp (http://www.amazon.co.jp/)
- 2ch.net (http://www.2ch.net/)

がある．Amazon.co.jp は大手オンラインブックストアの一つであり, 各本のページには **Amazon.co.jp ランキング**というものがある．このランキングは1時間に一度更新され, その本が購入されるとランキングは上がる．2ch.net は日本屈指のアクセス数を持つ電子掲示板の集合体であり, 掲示板を**板** (**いた**) と呼ばれるジャンルで分類している．さらに各板にはその分野に属する話題ごとに細かく分けられた**スレッド**と呼ばれる掲示板が 200〜750 存在する．また各板には**スレッドの一覧表**があり, スレッドに**書き込み**がされるとそのスレッドは一覧表の先頭にジャンプする[1]．例えば 2ch.net にアクセスし, ある板のスレッド一覧表が図 3.1 であったとする．

1:高校生のための数学の質問スレ PART215(177)　2:予選》╋|||《数学オリンピック 14》╋|||《近し (484)　3:分からない問題はここに書いてね２９９ (50)　4:ポケットのビスケットを叩いた場合二つに割れる確率 (26)　5:＼(^o^)／謹賀新年@数学板＼(^o^)／(31)　6:【激しく】解析と線型代数の本何がいい？【既出】(735)　7:雑談はここに書け！【３３】(703)　8:TeX 総合スレッド VII(737)　9:名古屋大学】 多元数理科学研究科 [Chapter 33](581)

⋮略

705:数字と数学どっちが先に生まれたの？(82)　706:数学のあるある (30)

図 3.1: スレッド一覧表の例. ランキング1から順にスレッドが並んでおり, 各スレッドの項目は ランキング, スレッド名, 書き込み数 の順である

この後スレッド "TeX 総合スレッド VII" に書き込みがされるとスレッド一覧表は図 3.2 になる．

[1] 2ch.net には書き込みをしても一覧表の先頭にジャンプさせない sage というものがあるが, ここでは sage による書き込みは無視することにする．

1:TeX 総合スレッド VII(738)　2:高校生のための数学の質問スレ PART215(177)　3:予選》╋|||《数学オリンピック 14》╋|||《近し(484)　4:分からない問題はここに書いてね２９９(50)　5:ポケットのビスケットを叩いた場合二つに割れる確率(26)　6:＼(^o^)／謹賀新年@数学板＼(^o^)／(31)　7:【激しく】解析と線型代数の本何がいい？【既出】(735)　8:雑談はここに書け！【３３】(703)　9:名古屋大学】 多元数理科学研究科 [Chapter 33](581)

⋮略

705:数字と数学どっちが先に生まれたの？(82)　706:数学のあるある(30)

図 3.2: スレッド"Tex 総合スレッド VII"に書き込みがあったため, 一覧表の一番上にジャンプした

この章では 2ch.net のランキングの時間発展に着目することにする.

## 3.1 オンラインランキングへの応用

2 章の理論をオンラインランキングへ応用するにあたり, ジャンプ率の分布　を仮定する必要がある. 様々な社会ランキングに対する基本的なモデルの分布として伝統的にパレート分布が使われる. もっとも良く知られた例は収入のランキングである. ここでもジャンプ率の分布　はパレート分布であると仮定する.

**定義 3.1.** $a, b$ を正の定数とする. 分布関数 $F(w)$ が

$$F(w) = \begin{cases} 1 - \left(\frac{a}{w}\right)^b; & w \geqq a \\ 0; & w < a \end{cases} \tag{3.1.1}$$

となる分布を**パレート分布**と云う. パレート分布の確率密度関数 $f(w)$ は

$$f(w) = \begin{cases} ba^b w^{-(b+1)}; & w \geqq a \\ 0; & w < a \end{cases}$$

となる.

ここでは離散の場合を考えることにする. $i = 1; 2; \cdots; N$ に対し, $i$ 番目に大きいジャンプ率を $w_i$ とする. (3.1.1) より

$$([w; \infty)) = \frac{1}{N} \#\{i \mid w_i \geqq w\} \sim \left(\frac{a}{w}\right)^b; \quad w \geqq a$$

であるから

$$w_i = a\left(\frac{N}{i}\right)^{\frac{1}{b}}; \quad i = 1; 2; \cdots; N \tag{3.1.2}$$

を得る. ゆえに次を仮定する.

**仮定 3.2.** $i = 1; 2; \cdots; N$ とし, $i$ 番目に大きいジャンプ率を $w_i$ とする. このとき

$$w_i = a\left(\frac{N}{i}\right)^{\frac{1}{b}}$$

であるとする.

**注 3.3.** パレート分布のパラメータ $a$ はジャンプ率が最も低い粒子のジャンプ率を表わしている. 実際に (3.1.2) において $w_N = a$ である. パラメータ $b$ はジャンプ率の不平等さを反映する. 実際に (3.1.2) において $w_1$=$w_N = N^{1/b}$ で, $b$ が大きければこの比は 1 に近く, $b$ が小さければこの比は大きい値になる.

(2.2.3) で定義される $y_C(t)$ は, 初期時刻にランキング 1 の粒子が時刻 $t$ において tail side にいるのなら, この粒子の時刻 $t$ でのランキングを表わしていた. 今仮定 3.2 より

$$\begin{aligned} y_C(t) &= 1 - ba^b \int_a^\infty e^{-wt} w^{-b-1} dw \\ &= 1 - b(at)^b \Gamma(-b; at) \end{aligned} \tag{3.1.3}$$

となる. ただし $\Gamma(z; p)$ は不完全ガンマ関数である. すなわち $z \in \mathbb{R}$, $p > 0$ に対し,

$$\Gamma(z; p) := \int_p^\infty e^{-w} w^{z-1} dw$$

である.

**注 3.4.** 任意の $b > 0$ に対して $\Gamma(-b; at)$ は発散してしまうが, (3.1.3) においては $t^b$ でキャンセルさせるため問題はない. なお (3.1.3) を部分積分することで,

$$y_C(t) = 1 - e^{-at} + (at)^b \Gamma(1 - b; at) \tag{3.1.4}$$

を得る. 従って $0 < b < 1$ のとき,

$$y_C(t) = a^b \Gamma(1 - b; 0) t^b + o(t^b)$$

である. すなわち $t = 0$ の近くでは $y_C(t)$ は上に凸で $t = 0$ における接線は垂直軸になる. なお, もう一度 (3.1.3) を部分積分することで

$$y_C(t) = 1 - \left(1 - \frac{at}{b-1}\right) e^{-at} - \frac{(at)^b}{b-1} \Gamma(2 - b; at)$$

を得る. 従って $1 < b < 2$ のとき,

$$y_C(t) = \frac{ab}{b-1} t - \Gamma(2 - b; 0) \frac{a^b}{b-1} t^b + O(t^2)$$

である. すなわち $t = 0$ の近くでは $y_C(t)$ は線型になる.

## 3.2 オンラインランキングの分析

(3.1.4) より $x_C(t) \sim N y_C(t)$ は,

$$x_C(t) = N\left\{1 - e^{-at} + (at)^b \Gamma(1-b, at)\right\} \tag{3.2.1}$$

となる. ゆえに $x_C(t)$ のグラフは図 3.3 のような一本の曲線になる.

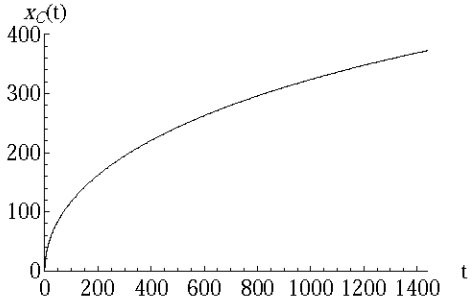


図 3.3: $N = 700$, $a = 10^{-4}$, $b = 0.50$ としたときの $x_C(t)$ のグラフ

オンラインランキングの分析を行うために, 指定した一つの板のスレッド一覧表でランキング変動があるごとにその時刻とランキングを自動で収集するプログラムを作成した[2]. 図 3.4 はある板のスレッド一覧表でのランキング変動を某日 24 時間自動収集して, そのデータをプロットしたものである. なおスレッド総数 $N$ は $N = 697$ である.

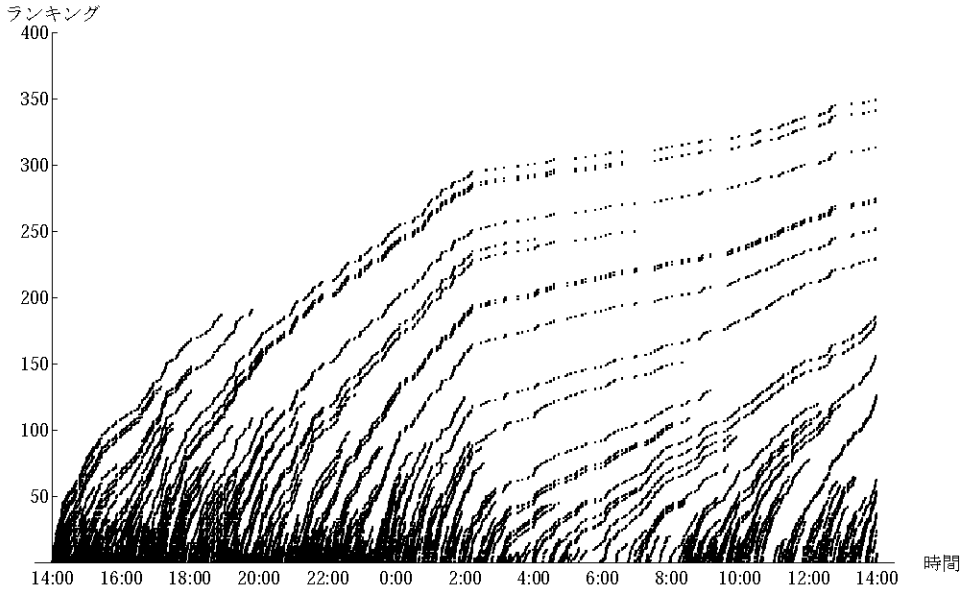


図 3.4: スレッド一覧表で, ランキング変動があったときの時刻と変動後のランキングのプロット図　観測日時：2008 年 11 月 22 日 14 時から 2008 年 11 月 23 日 14 時

[2]詳細は付録 A で述べるがここでは言語 Perl を用いた.

図 3.4 においてプロットを, ランキング 1 になってから次にランキング 1 になる手前までを一つの区切りであると考えると, その区切ったもの一つ一つはまさに $x_C(t)$ である. 図 3.5 は図 3.4 から幾つか抜き出したものである.

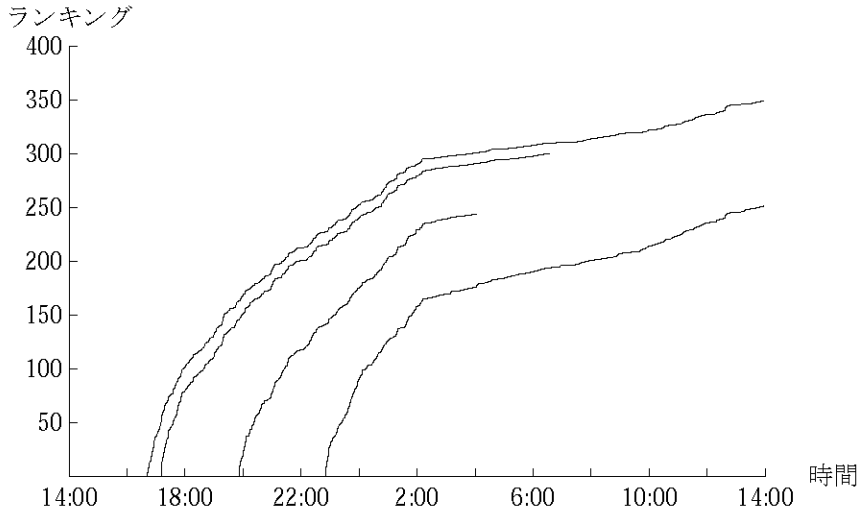


図 3.5: ランキング 1 になってから次にランキング 1 になる手前までのプロットを結んだグラフの例

図 3.6 は 24 時間で得られた曲線 $x_C(t)$ のグラフをそれぞれランキング 1 の時刻を 0 となるように平行移動したものである. なおスレッド総数は $N = 697$ であった.

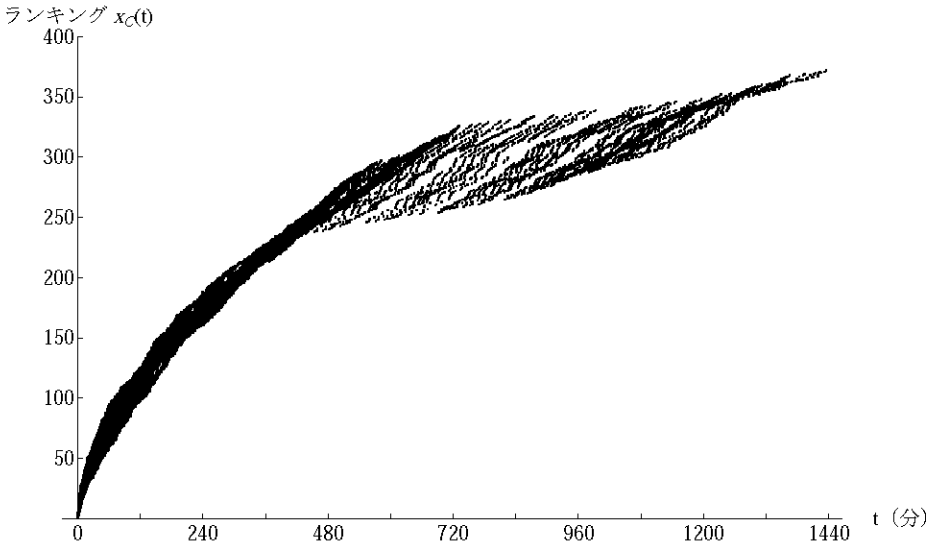


図 3.6: 2008 年 11 月 22 日 14 時から 2008 年 11 月 23 日 14 時までの 24 時間で得られた $x_C(t)$ のグラフをそれぞれランキング 1 の時刻を 0 となるように平行移動したもの. データ総数 $n_d = 26,147$

このことを用いて2ch.netのジャンプ率 (書き込み率) の分布 $\lambda$ のパラメータ $(a,b)$ を最小二乗法を用いて推定する. すなわち $(t_i,x_i)$ を $i$ 個目のデータ ($x_i$ はランキング, $t_i$ は時刻), 時刻 $t$ での理論値を $f_{a,b}(t)$, データ総数を $n_d$, とし,

$$E(a,b) := \sum_{i=1}^{n_d} \{x_i - f_{a,b}(t_i)\}^2$$

を最小にする $(a,b)$ を求める. ただしここでは粒子総数 (スレッド総数) $N$ は変動しないものとして考えることにする. (3.2.1) より,

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

である. Mathematicaを用いて $(a,b)$ の推定値 $(a^*,b^*)$, 残差変動 $E_{\min} := E(a^*,b^*)$, 決定係数 $R := \sqrt{1 - \dfrac{E_{\min}}{S_{\text{total}}}}$ を求める. ただし,

$$S_{\text{total}} := \sum_{i=1}^{n_d} (x_i - \bar{x})^2$$

で定義し, 全変動と云う. なお $\bar{x}$ は標本平均である. 決定係数 $R$ は標本値から求めた回帰曲線のあてはまりの良さの尺度であり, 1に近いほど良くあてはまっているといえる.

**24時間のデータより求めた値**

$$a^* = 1.27953 \times 10^{-4}, \quad b^* = 0.5600,$$
$$E_{\min} = 5.601690 \times 10^6, \quad \sqrt{\frac{E_{\min}}{n_d}} = 14.6369, \quad R = 0.988366.$$

次の図3.7は図3.6のデータに回帰曲線をあてはめたものである.

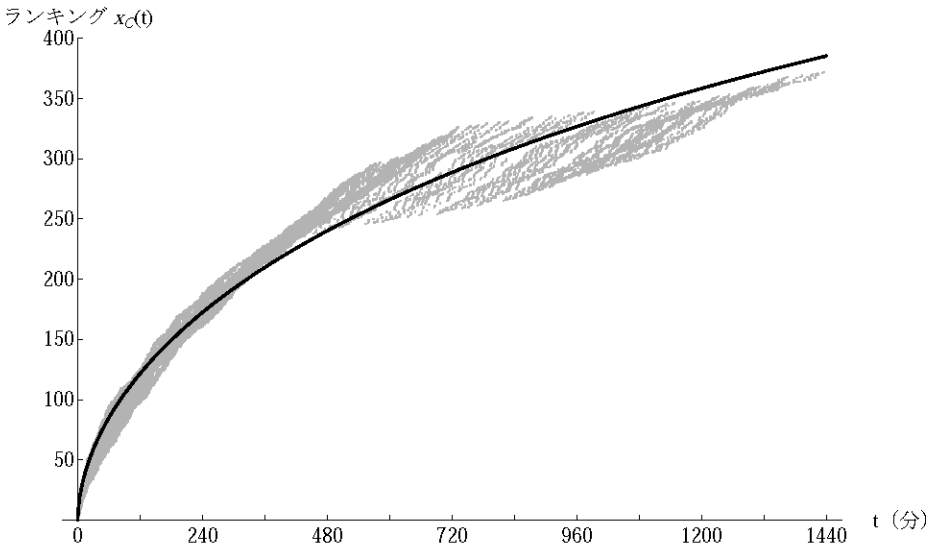


図 3.7: 図3.6と理論曲線

## 3.3 ジャンプ率の時間依存性

前節では 24 時間のデータを基に 2ch.net のランキングの時間発展の様子を分析した. しかし図 3.4, 図 3.6 を見るとよく分かるが, 24 時間においてはジャンプ率は一定ではなく, 昼夜でジャンプ率の差があると考えるのが尤もらしいといえるので, 2 章の内容に時間依存性を入れる拡張をする. 拡張するにあたり粒子 $i$ の第 $j$ ジャンプ時刻 $\tau_{i,j}^{(N)} : \Omega \to \mathbb{R}_+$ について, 仮定 2.3 の代わりに次を仮定する.

**仮定 3.5.** $i = 1; 2; \cdots ; N$ に対し, $w_i^{(N)}(t) : \mathbb{R}_+ \to (0; \infty) \in D$ とする. ただし, $D$ は *cádlág* 関数全体の集合, すなわち

$$D = \{ f(t) : \mathbb{R}_+ \to (0; \infty) \mid \text{右連続; 左極限が存在する} \}$$

であるとする. このとき,

(1) $\tau_{i,0}^{(N)} = 0$,

(2) $\tau_i^{(N)}$, $i = 1; 2; \cdots ; N$ は独立,

(3) $\tau_i^{(N)}$ の法則は $t \geqq 0$ に対し

$$\mathbf{P}[\tau_i^{(N)} \leqq t] = 1 - \exp\left\{ - \int_0^t w_i^{(N)}(s) ds \right\} \tag{3.3.1}$$

であるとする.

$w_i^{(N)}(t)$ は粒子 $i$ の時刻 $t$ におけるジャンプ率を表わす. ジャンプ率 $\{w_i(\cdot)\}_{i=1}^N$ の経験分布を $\Lambda^N$ とおく. すなわち $\Lambda^N$ は, $\Lambda^N$ に関する期待値を $\mathbf{E}^{\Lambda^N}$ と書くことにすると,

$$\mathbf{E}^{\Lambda^N}[g(w)] = \frac{1}{N} \sum_{i=1}^N g(w_i^{(N)})$$

を満たすものとする. このときも 定理 2.10 と同様に, $N \to \infty$ で $y_C^{(N)}(t)$ は決定論的極限に収束することが従う.

**定理 3.6.** $\Lambda^N$ は $N \to \infty$ のとき, 関数空間 $D$ 上のある分布 $\Lambda$ に弱収束しているとする. すなわち任意の $g \in C_b(D; \mathbb{R})$ に対し,

$$\lim_{N \to \infty} \mathbf{E}^{\Lambda^N}[g(w)] = \mathbf{E}^{\Lambda}[g(w)] \tag{3.3.2}$$

であるとする. このとき

$$\lim_{N \to \infty} y_C^{(N)}(t) = y_C(t) \quad \textit{in prob:}$$

が従う. ただし

$$y_C(t) = 1 - \mathbf{E}^{\Lambda}\left[\exp\left\{-\int_0^t w(s)ds\right\}\right] \tag{3.3.3}$$

である.

**証明**.

$$y_C^{(N)}(t) = \frac{1}{N}\sum_{i=1}^{N}\left\{ {}_{\tau_i^{(N)} \leqq t} - \mathbf{P}[{}_i^{(N)} \leqq t]\right\} + \frac{1}{N}\sum_{i=1}^{N}\mathbf{P}[{}_i^{(N)} \leqq t]$$

である. 仮定 3.5 より ${}_{\tau_i^{(N)} \leqq t}$, $i = 1; 2; \cdots; N$ は独立である. また (3.3.1) より

$$\begin{aligned}\mathbf{E}[{}^2_{\tau_i^{(N)} \leqq t}] &= \mathbf{P}[{}_i^{(N)} \leqq t] \\ &= 1 - \exp\left\{-\int_0^t w_i^{(N)}(s)ds\right\} \\ &< 1\end{aligned}$$

が従うので, 補題 2.11 より

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left\{ {}_{\tau_i^{(N)} \leqq t} - \mathbf{P}[{}_i^{(N)} \leqq t]\right\} = 0 \quad \textit{in prob:}$$

が従う. また (3.3.2), (3.3.3) より

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{P}[{}_i^{(N)} \leqq t] &= 1 - \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\exp\left\{-\int_0^t w_i^{(N)}(s)ds\right\} \\ &= 1 - \lim_{N\to\infty}\mathbf{E}^{\Lambda^N}\left[\exp\left\{-\int_0^t w(s)ds\right\}\right] \\ &= 1 - \mathbf{E}^{\Lambda}\left[\exp\left\{-\int_0^t w(s)ds\right\}\right] \\ &= y_C(t)\end{aligned}$$

が従う. 以上のことから定理 3.6 を得る. □

この拡張に伴いジャンプ率の分布は次に従うと仮定する.

**仮定 3.7.** ジャンプ率の分布はパラメータが時間変動するパレート分布, すなわち $i = 1; 2; \cdots; N$, $t \geqq 0$ に対し,

$$w_i(t) = a(t)\left(\frac{N}{i}\right)^{\frac{1}{b}}$$

であるとする. ただし $b$ は正定数, $a : \mathbb{R}_+ \to (0; \infty)$ は単関数であるとする.