# 確率的順位付け

服部哲弥　（慶應義塾大学・経済）

2011.07.19–22 集中講義（京都大学理学部）

# 1．確率的順位付け模型

$N$ 自然数（粒子数）

$X^{(N)} = (X_1^{(N)}, \cdots, X_N^{(N)}) : \ \Omega \times \mathbb{R}_+ \to \{1, 2, \ldots, N\}^N$

（ $1, 2, \cdots, N$ の順列に値をとる確率過程 $X^{(N)} = X^{(N)}(t)$ ）

初期値： $X_i^{(N)}(0) = x_i^{(N)}, \ i = 1, 2, \cdots, N$；

$\{x_1^{(N)}, x_2^{(N)}, \ldots, x_N^{(N)}\} = \{1, 2, \ldots, N\}$

時間発展：サンプル毎にポワッソンのランダム測度に応じた

（有限時間内では有限個の）時刻でのみ $X^{(N)}(t)$ は変化（続）

Stochastic ranking process
服部K，服部T, SPA (2009) 一様
服部K，服部T, FE (2009) 極限PDE
針谷, 服部K, 服部T, 永幡, 竹島, 小林, TMJ (2011) 時刻依存性
永幡 (2011) 分布値確率過程の収束
服部T, 楠岡 (2011) 位置依存性

# ポワッソン過程（ポワッソンのランダム測度）

**添え字 $N, i$ を一時的に外す**

$w: \ \mathbb{R}_+ \to \mathbb{R}_+$: **ジャンプ率関数**，　$\rho((0,t]) = \int_0^t w(s)\,ds$

$\nu: \ (\Omega, \mathcal{B}(\mathbb{R}_+)) \to \mathbb{Z}_+$

**（増分１の）強度 $\mathrm{E}[\ \nu\ ] = \rho$ のポワッソンランダム測度，すなわち，**

- **確率１のサンプル $\omega \in \Omega$ 毎に $\nu(\omega, \cdot): \ \mathcal{B}(\mathbb{R}_+) \to \mathbb{Z}_+$ は測度**

$\boxed{\text{（}\mathbb{R}_+\text{ にばらまかれた「個数」）}}$

- $\nu(A) = \nu(\cdot, A)$ **は平均 $\rho(A)$ のポワッソン分布に従う確率変数**
- $A \cap B = \emptyset$ **ならば $\nu(A)$ と $\nu(B)$ は独立**

**例．** $\rho((0,t]) = t$ **のとき，** $Z_t = \nu((0,t])$ **は一様ポワッソン過程（昼夜差無し）**

# 時間発展

$N$ 固定．$i=1,2,\ldots,N$

$\rho_i^{(N)}:\ \mathcal{B}(\mathbb{R}_+)\to\mathbb{R}_+$ （連続な測度）

$\nu_i^{(N)}:\ (\Omega,\mathcal{B}(\mathbb{R}_+))\to\mathbb{Z}_+$ $\rho_i^{(N)}=\mathrm{E}[\,\nu_i^{(N)}\,]$ を強度とする増分 1 のポワッソンランダム測度；

$i$ について独立（$\{\nu_i^{(N)}(A_i)\mid i=1,2,\ldots,N\}\perp$），

$$X_i^{(N)}(t)=x_i^{(N)}+\sum_{k=1}^{N}\int_0^t \mathbf{1}_{X_k^{(N)}(s-0)>X_i^{(N)}(s-0)}\ \nu_k^{(N)}(ds)$$

$$+\int_0^t (1-X_i^{(N)}(s-0))\,\nu_i^{(N)}(ds),\ i=1,\cdots,N,\ t\geqq 0$$

$\mathbf{1}_A$ は事象 $A$ の定義関数

$X_i^{(N)}(t)$： 時刻 $t$ の粒子 $i$ の順位

# 先頭に跳ぶ規則

**（再掲）** $X_i^{(N)}(t) = x_i^{(N)} + \sum_{k=1}^{N} \int_0^t \mathbf{1}_{X_k^{(N)}(s-0) > X_i^{(N)}(s-0)} \, \nu_k^{(N)}(ds)$

$$+ \int_0^t \left(1 - X_i^{(N)}(s-0)\right) \nu_i^{(N)}(ds)$$

**各積分は $\nu_i^{(N)}$ の増分が 1 となる時刻での被積分関数の値の和**

$J_i^{(N)}(s,t) := \{\nu_i^{(N)}((s,t]) > 0\} \in \mathcal{F}$

**時間発展の右辺第２項→**

(1) $J_i^{(N)}(s,t)$ **は粒子 $i$ が $(s,t]$ の間に 1 位に跳ぶ事象**

**時間発展の右辺第１項→**

(2) **粒子 $i$ より下位の粒子が 1 位に跳ぶ毎に 1 ずつ順位が下がる**

$X_i^{(N)}(t)$ **＝最後に跳んだ順に並べるときの粒子 $i$ の位置（順位）**

# Move-to-front rule

M.L.Tsetlin (1963) **定常分布（$N$ 個の順列上の確率過程として）**
J.McCabe (1965) **探索コスト（先頭に跳ぶ直前の位置の期待値）**
1970 **年代再発見** J.A.Fill **無限粒子極限（個別のジャンプ率分布）**
1990–2000 **年代** P.R.Jelenković **（個別のジャンプ率分布，拡張）**

**新しいこと**

- **位置－強度（ジャンプ率）結合経験分布の大数の法則**
  - **偏微分方程式で特徴づけられる分布への収束（流体力学的極限）**
- Amazon **ランキングの「ロングテール分析」**
  - **流行に従う順位付けの究極の単純化の視点（超整理法，積ん読）**

# ＊．目次

## 2． 大数の法則（特性曲線，結合経験分布）

（再掲）粒子 $i$ が時刻 $t$ までに先頭に跳ぶ事象 $J_i^{(N)}(0,t) = \{\nu_i^{(N)}((0,t]) > 0\}$

**特性曲線**（の，離散版） $Y_C^{(N)}(t) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \mathbf{1}_{J_i^{(N)}(0,t)}$

**ジャンプ済み粒子と未ジャンプ粒子の境界**

**命題．** $\lambda_t^{(N)} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \delta_{\rho_i^{(N)}((0,t])} \to \exists\lambda_t \ (N \to \infty)$ ならば，

$$Y_C^{(N)}(t) \to y_C(t) = 1 - \int_0^\infty e^{-s}\lambda_t(ds) \ (N \to \infty，概収束) \qquad \diamond$$

**証明：** 独立確率変数の和の大数の強法則：$Y_C^{(N)}(t) - \mathrm{E}[\, Y_C^{(N)}(t)\,] \to 0$

＋ポワッソン分布：$\mathrm{P}[\, J_i^{(N)}(0,t)^c\,] = \mathrm{P}[\, \nu_i^{(N)}((0,t]) = 0\,] = e^{-\rho((0,t])}$

# 位置－強度結合経験分布の収束

**強度 $\rho$ と規格化順位 $Y_i^{(N)} = \frac{1}{N}(X_i^{(N)} - 1)$ の結合経験分布：**

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

**定理．** $\mu_0^{(N)} \to \exists \mu_0 \ (N \to \infty)$，かつ，各 $0 \leqq s < t$ に対して $N \to \infty$ で

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{\rho_i^{(N)}}((s,t]) \to \int_{\mathcal{M}(\mathbb{R}_+)} \delta_{\rho((s,t])} \Lambda(d\rho); \qquad \Lambda(d\rho) = \mu_0(d\rho \times [0,1))$$

ならば，各 $t > 0$ に対して $\mu_t^{(N)} \to \mu_t \ (N \to \infty$，概収束）．
$\mu_t$ は（初期分布 $\mu_0$ を用いてあらわに書ける）非ランダムな分布． $\diamond$

# 極限を記述する偏微分方程式

$U(d\rho, y, t) := \mu_t(d\rho \times [y, 1))$ (「分布関数」)

強度が有限種類で密度（ジャンプ率）を持つ場合：

$$\Lambda = \sum_\beta r_\beta \delta_{\rho_\beta},\ \rho_\beta(A) = \int_A w_\beta(u)\, du,\ w_\beta, r_\beta > 0$$

**定理.** $\displaystyle\sum_\beta r_\beta = 1,\ \sum_\beta r_\beta w_\beta(t) < \infty,$

$u_\alpha:\ [0, 1) \to \mathbb{R}_+$: 非負滑らか狭義減少，$\displaystyle\sum_\beta u_\beta(y) = 1 - y$

ならば，$U_\alpha(y, t) = U(\{\rho_\alpha\}, y, t)$ は次の偏微分方程式系の初期値問題の時間大局的一意古典解：

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

境界条件 $U_\alpha(0, t) = r_\alpha$，初期値 $U_\alpha(\cdot, 0) = u_\alpha(\cdot)$, ◇

# 証明の鍵　ー　特性曲線$y_C(y_0, t_0; t)$

$\Gamma_i = \{(y,0) \in [0,1) \times \mathbb{R}_+ \mid y \geqq 0\}$（初期点集合），

$\Gamma_b = \{(0,t) \in [0,1) \times \mathbb{R}_+ \mid t \geqq 0\}$（境界点集合），$\Gamma = \Gamma_i \cup \Gamma_b$

$(y_0, t_0) \in \Gamma$ に $Y_C^{(N)}(t) = Y_C^{(N)}(y_0, t_0; t)$ と $y_C(t) = y_C(y_0, t_0; t)$ を拡張

$$Y_C^{(N)}(y_0, t_0; t) \;:= y_0 + \frac{1}{N} \sum_{i;\, Y_i^{(N)}(t_0) \geqq y_0} \mathbf{1}_{J_i^{(N)}(t_0, t)}$$

$$y_C(y_0, t_0; t) \;:= 1 - \int_{\mathcal{M}(\mathbb{R}_+)} e^{-\rho((t_0, t])} \mu_{t_0}(d\rho, [y_0, 1))$$

$(y_0, 0) \in \Gamma_i$ に対して $y_C(y_0, 0; t) = 1 - \int_{\mathcal{M}(\mathbb{R}_+)} e^{-\rho((0,t])} \mu_0(d\rho, [y_0, 1))$

$(0, t_0) \in \Gamma_b$ に対して $y_C(0, t_0; t) = 1 - \int_{\mathcal{M}(\mathbb{R}_+)} e^{-\rho((t_0,t])} \Lambda(d\rho)$

大数の法則（独立確率変数列）：$Y_C^{(N)}(y_0, t_0; t) \to y_C(y_0, t_0; t)$

# 証明の鍵　－　分布関数

**有限種類の場合：** $i \mapsto \alpha = \alpha(N,i);\ \rho_i^{(N)} = \rho_{\alpha(N,i)}$

**（再掲）位置に関する「分布関数」** $U_\alpha(y,t) = U(\{\rho_\alpha\},y,t) = \mu_t(\{\rho_\alpha\} \times [y,1))$

**時空１＋１次元：** $U_\alpha^{(N)}(y,t) = \mu_t^{(N)}(\{\rho_\alpha\} \times [y,1))$

$$= \frac{1}{N} \sum_{i;\,\alpha(N,i)=\alpha} \mathbf{1}_{X_i^{(N)}(t-) \geqq Ny+1} \quad \Leftrightarrow \quad \mu_t^{(N)}$$

**（再掲）** $Y_C^{(N)}(t) = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} 1_{J_i^{(N)}(0,t)}$ **：未ジャンプ粒子の境界**

- $U_\alpha^{(N)}(Y_C^{(N)}(t),t)$ **は先頭に跳ぶ項のみで時間発展**（$N < \infty$）
- **主部が共通の１階準線形ＰＤＥ→特性曲線の方法**（$N = \infty$）

# ３．データの統計的当てはめ（方法）

（再掲）$X_i^{(N)}(t)$ ＝最後に跳んだ順に並べるときの粒子 $i$ の順位

**流行度に応じた順位の，もっとも簡単な確率モデル**：**購入毎に１位**

流行度を反映する大規模逐次更新順位データ（ランキング）に適用可能

$x_i^{(N)} = X_i^{(N)}(0) = 1$ のとき，先頭に跳ぶまで（事象 $J_i^{(N)}(0,t)^c$ が成り立つ $t$ で）

$$X_i^{(N)}(t) = X_C^{(N)}(t) + 1$$

インターネットの日常化で**現実的可能性**（例：オンラインリテール）

- インターネット書店（通販サイト） Amazon.co.jp の本のランキング（Amazon.com の sales rank）
- 巨大掲示板集合体 2ch.net のスレッド一覧

（可能性：ブログサイト閲覧順位，学術論文の機関リポジトリでの参照順位）

# 流れの中の粒子の軌道

**（再掲）　命題．** $\lambda_t^{(N)} \to \lambda_t$ **ならば，** $Y_C^{(N)}(t) \to y_C(t) = 1 - \int_0^\infty e^{-s}\lambda_t(ds)$

$$X_i^{(N)}(t) \sim N\,Y_C^{(N)}(t) \sim N(1 - \int_0^\infty e^{-s}\lambda_t(ds)) \sim \sum_{i=1}^{N}(1 - e^{-\rho_i^{(N)}((0,t])})$$

(on $J_i^{(N)}(0,t)^c$)

**逆ラプラス変換が強度** $\rho_i^{(N)}((0,t])$ **の（$i$ についての）分布**

**有限個のデータで強度関数（無限パラメータ）は原理的に定まらない**

**限られたデータによって，社会についての何を知るべきか？**

# ３－１．昼夜差

- **共通の時刻依存性の仮定**：$\rho_i^{(N)}((0,t]) = \ w_i^{(N)} A(t),\ w_i^{(N)} > 0$

**社会活動は昼夜差 がある　（先取り：実際のデータで無視できない）**
**昼夜差は順位の対象に無関係な共通の（強度の）時刻依存性と仮定**
**（差は統計誤差，と扱う単純化）**

$$\lambda_t \sim \lambda_t^{(N)} = \frac{1}{N}\sum_{i=1}^{N} \delta_{w_i^{(N)} A(t)}$$ **における時刻依存性** $(A(t))$ **と粒子差**

$(w_i^{(N)})$ **の分離**

## 累積ジャンプ数

**時刻依存性 $A(t)$ についてのデータ処理**

**累積総ジャンプ数** $S_n^{(N)}(t) = \sum_{i=1}^{n} \nu_i^{(N)}((0,t])$ **で時間を計る**

**大数の法則：**

$$S_n^{(N)}(t) \sim \sum_{i=1}^{n} \rho_i^{(N)}((0,t]) = A(t)Z(N,n);\ Z(N,n) = \sum_{i=1}^{n} w_i^{(N)}$$

**（定数倍を積分変数に吸収すれば）** $A(t) = S_n^{(N)}(t)$

**以下，簡単のため，** $n = N,\ Z(N) = Z(N,N),\ S_n^{(N)} = S^{(N)}$

# ジャンプ率分布

$$X_i^{(N)}(t) \sim \sum_{i=1}^{N}(1 - e^{-\rho_i^{(N)}((0,t])}) \sim N(1 - \int_0^\infty e^{-wS^{(N)}(t)}\lambda(dw));$$

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

**強度の時刻依存性（昼夜差）を対象に無関係に共通と仮定することで，統計的当てはめが単純化：**

$A(t) \quad \leftarrow \quad S^{(N)}(t)$ **時刻依存性**

$\lambda \quad \leftarrow \quad X_i^{(N)}(t)$ **（相対）ジャンプ率分布**

# 日内変動に影響されないデータ採取計画法

時刻の関数 $A(t)$ を決める $(S^{(N)}(t))$ には大規模なデータが必要

ジャンプ率分布 $\lambda$ を先に引き出したいとき

（時刻依存性は日内変動としたので，積分変数を取り直せば）$A(t) = t+$ 周期関数

$$X_i^{(N)}(24n + t_0) \sim N(1 - \int_0^\infty e^{-wn+a_0}\lambda(dw))$$

- 毎日定時に得たデータのみ用る（逃したら諦める）場合は，日内変動が無い場合に一致

# ３－２．Pareto分布

（相対）ジャンプ率の分布 $\lambda$

（再掲）$X_i^{(N)}(t) \sim N(1-\int_0^\infty e^{-wS^{(N)}(t)}\lambda(dw))$

または $X_i^{(N)}(24n+t_0) \sim N(1-\int_0^\infty e^{-wn+a_0}\lambda(dw))$

有限個のデータ→ラプラス逆変換を厳密には取れない→

社会的特徴を引き出すよう単純化

社会学などでは（一般化）Zipfの法則(連続版はPareto分布)がよく用いられる

$w_i^{(N)} = a\left(\frac{N}{i}\right)^{1/b}$；$a$: 最低収入，$b$: 平等性の指数

$S^{(N)}(t)=t$（一様強度）のとき，

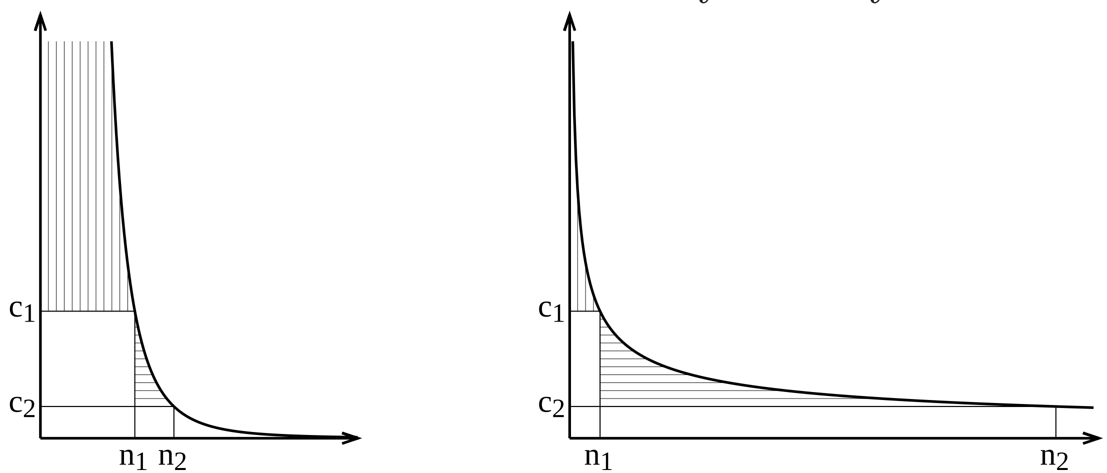
$X_i^{(N)}(t) \sim N(1-b(at)^b\Gamma(-b,at))$; $\Gamma(z,p)=\int_p^\infty e^{-x}x^{z-1}\,dx$

$N,a,b$を与えれば決まる（データを使って統計的当てはめ）

# 指数$b$

多品目商品を扱う小売業（例：大型書店）　$X_i^{(N)}$ ⇔ランキングデータ

$i$ **⇔**商品（本のタイトル），$w_i^{(N)}$ $(\rho_i^{(N)})$ **⇔**単位時間当売上，




縦軸⇔$w_i^{(N)}$，横軸⇔$i$，

左図：$b < 1$ 格差大　右図：$b > 1$ 平等に近い

$b > 1$ のとき，低コスト$(c_2 \ll c_1)$ならば，非売れ筋の売上**合計**が無視できない（**ロングテール**ビジネスモデル成立の可能性）　（後述）

# パラメータ $N$ と $a$

## パラメータの意味

（再掲） $b$：平等性（$b>1$ ⇔ロングテールビジネス成立可能性）

$N$：実効総冊数（Amazon.co.jp は絶版本も順位が付いている）
「パレート＋非売品」:「公称」総冊数 $N_0$，

$\lambda([w,\infty)) = (1-c)\left(\frac{a}{w}\right)^b,\ w \geqq a,$ かつ $\lambda(\{0\}) = c$ のとき

$$X_i^{(N)}(t) \sim N_0(1-\int_0^\infty e^{-wt}\lambda(dw)) \quad \sim N_0(1-c)(1-b(at)^b\Gamma(-b,at))$$

→ $N = N_0(1-c)$ がパラメータとして得られる

$a$：ロングテール側の切断（$b>1$ のとき必要）

（ビッグヒット支配的 $(b<1)$ のときは離散性が売上有限を保証）

# ３－３．主定理の極限分布

時間的に一様 $(\rho_i^{(N)}((0,t]) = w_i^{(N)} t)$，かつ，定常 $(t \to \infty)$ な場合
主定理の極限結合分布は

$$\mu_\infty(dw \times [y,1)) = e^{-wt_0(y)}\lambda(dw);\ 1 - y = \int_0^\infty e^{-wt_0(y)}\lambda(dw)$$

特に，$t_0(y) > 0,\ y > 0 \quad (\mathrm{cf.}\ \ t_0(0) = 0)$

時刻 $t$ 以降最初に跳ぶ粒子 $I(t)$ の直前位置 $C_N(t) = X^{(N)}_{I^{(N)}(t)}(t)$

$$\mathrm{P}[\ \frac{1}{N} C_N(t) \geqq y\ ] \to \frac{\int_0^\infty w\, \mu_\infty(dw \times [y,1))}{\int_0^\infty w\, \lambda(dw)}$$

先頭に跳ぶ粒子のうち tail 側 $[y,1)$ にいたものの割合（大数の法則）

# ランキング下位からの売上への寄与

（再掲）

(*) 先頭に跳ぶ粒子のうちtail側 $[y, 1)$ にいたものの割合 $\sim \dfrac{\int_0^\infty w\,\mu_\infty(dw \times [y,1))}{\int_0^\infty w\,\lambda(dw)}$

(**) $t_0(y) > 0,\ y > 0$

(***) 流行度に応じた順位の，もっとも簡単な確率モデル：購入毎に１位

(***)→　オンライン小売業では (*) は下位からの売り上げへの寄与

(**)　→　(*) の分子は有限　→

$\lambda$ の平均が有限⇔下位の売上は無視できない　　$(b < 1)$

$\lambda$ が平均を持たない⇔下位の売上は無視できる　　$(b > 1)$

# ロングテール

**平均売上頻度** $w_i^{(N)}$ **が下位の商品** $i$ **の平均売上への寄与**

$$\sum_{i;\, w_i^{(N)} \leqq w_0} w_i^{(N)} \sim N \int_0^{w_0} w\, \lambda(dw) < \infty$$

$\int_0^\infty w\, \lambda(dw) = \infty$ **⇔ビッグヒット支配的**　$(b < 1)$

$\int_0^\infty w\, \lambda(dw) < \infty$ **⇔ロングテールビジネス可能性**　$(b > 1)$

**下位の** $w_i^{(N)}$ **は（データ収集の意味で）測定可能量ではない**

**特に流行に関しては，迅速な経営判断が必要**

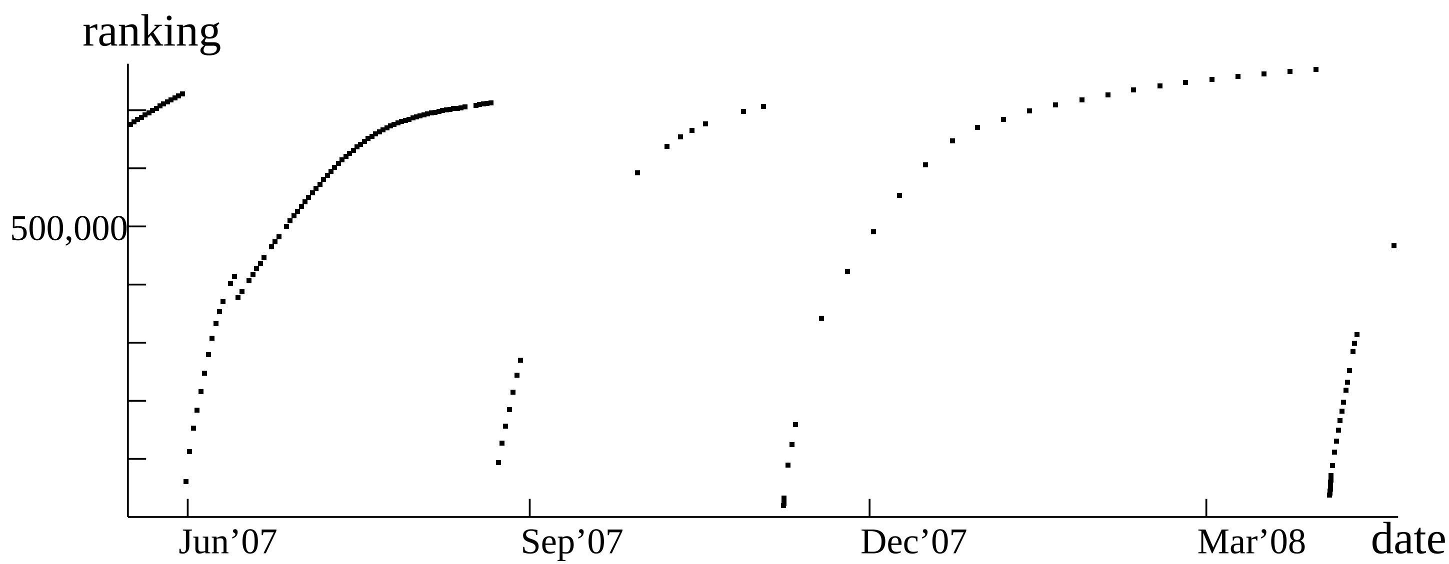
**（損失覚悟で）ある瞬間のランキングの下位を切る**（入学試験の原理）

**（機会）損失の見積もり**　（再掲）

- $\lambda$ **の平均が有限⇔下位の売上は無視できない**　$(b < 1)$
- $\lambda$ **が平均を持たない⇔下位の売上は無視できる**　$(b > 1)$

# ４．現実のデータ（指数の普遍性，強度の昼夜差）

Amazon.co.jp の（さほど売れない）本のランキングを**しつこく**観察



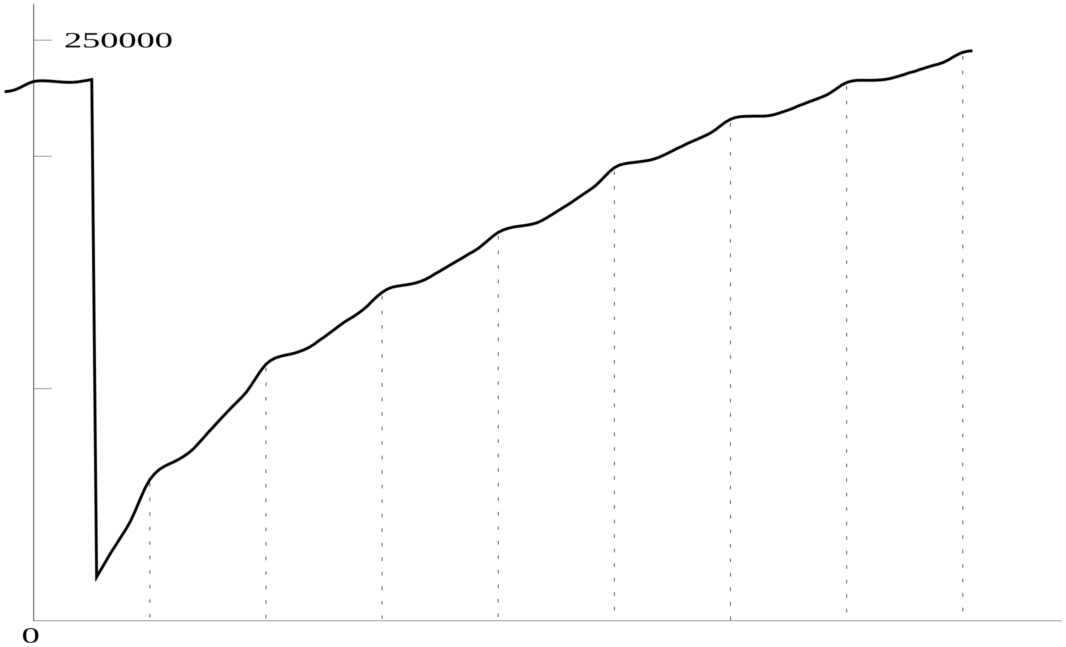

**上位への大きな跳び⇔アマゾン書店での注文（１冊でも）** →

（**アマゾン書店のランキングのアルゴリズムは公開されてないが,**）

**確率順位付け模型の結論を当てはめてみる**

$O$(１万) 位は１位とみなす近似で $O$(100 万) 位の現象を理解

# Amazon.co.jp　社会活動の昼夜差




明らかな昼夜差（深夜から早朝まで全体の活動停滞）
→強度の共通時刻依存性を適用
→毎日定時のデータは一様な場合に帰着

# アマゾン書店はロングテールではない！

**データの統計的当てはめ**

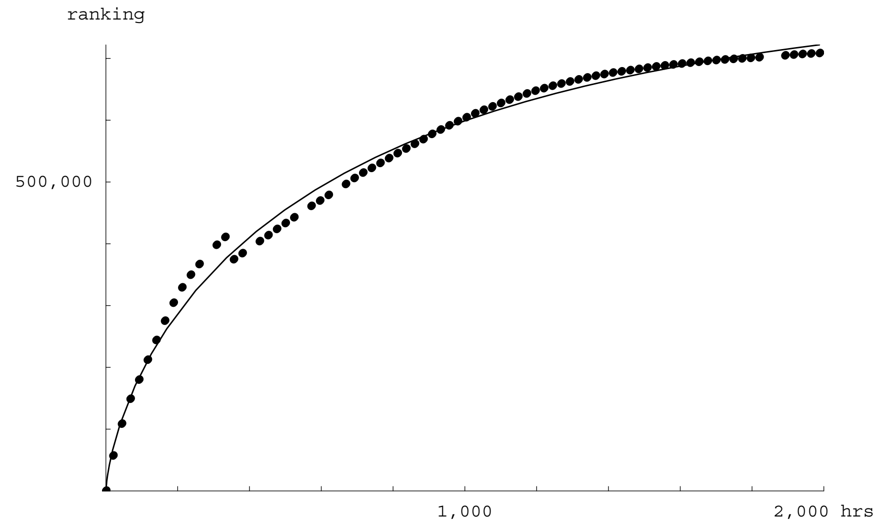
**（確率順位付け模型＋強度の共通時刻依存性＋注文頻度 Pareto 分布）**

**毎日 21 時のランキング**

’07.6–9 **（逃したら諦める）**

$(N^*,\ a^*,\ b^*) =$

$(8 \times 10^5,\ 5 \times 10^{-4},\ 0.77)$




- $b < 1$ **「不平等」上位** $o(N)$ **冊が売り上げの全てを占める**

**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**

## 「先行研究」との比較

Chevalier, Goolsbee $b = 1.2$ (Online bookstore の価格弾力性)

Brynjolfsson, Hu, Smith $b = 1.148$ (consumer welfareの評価)

多数の本の期間内購入数と順位を対応させる原始的（かつ，

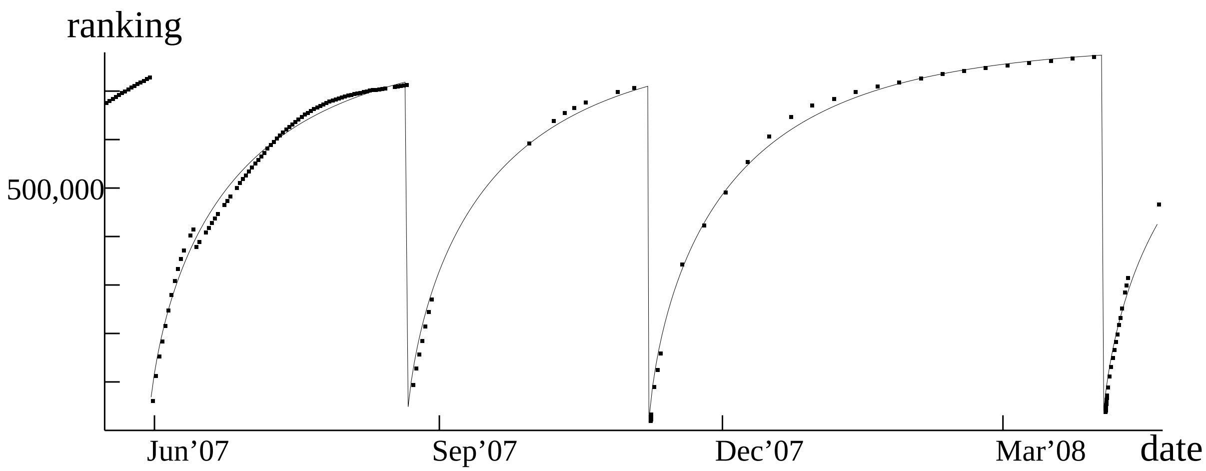
我々の視点からは誤った）方法．しかも，$b > 1$ は我々の結果と矛盾

Online retailの経済効果は先行主張ほどはない？

我々の数学的方法は社会学的（経済学的？）研究にも有効な可能性

# データの統計的当てはめの検証

- **相対購買頻度分布**




3パラメータで98点のデータを当てはめ

$$(N^*,\ a^*,\ b^*) = (8 \times 10^5,\ 6 \times 10^{-4},\ 0.81)$$

- **強度の時刻依存性**（昼夜差）のデータからの定量は難しい

アマゾン書店は1時間に一度の更新（**粗いデータ**）

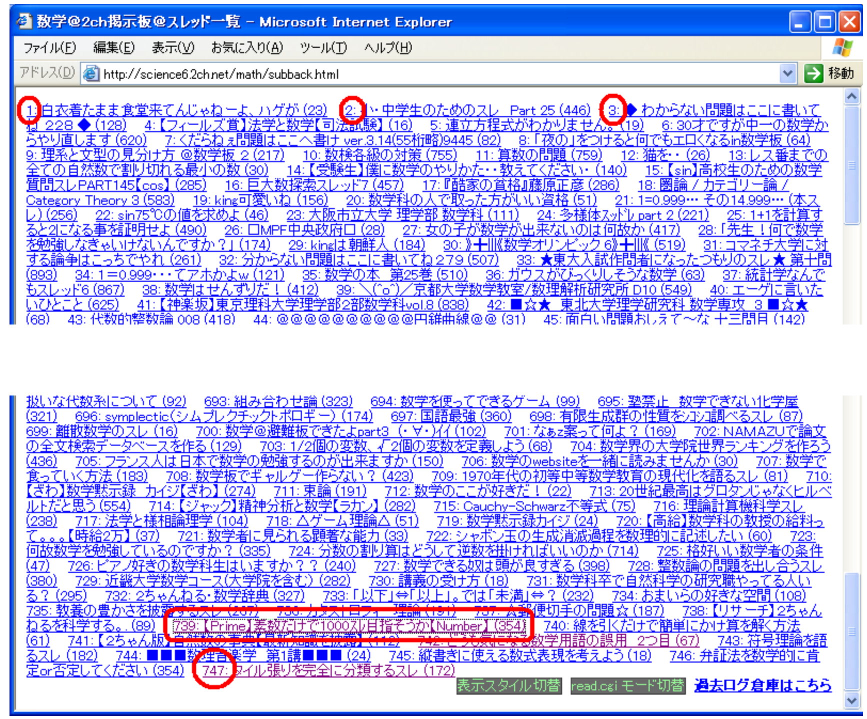
アマゾン書店のウェブ1ページに1点の本の情報

→強度（期待値）を得るには多数の本の累積注文数→サーバに負担

# 2ch.netのスレッド一覧

**２ちゃんねる：**
web**掲示板の巨大な集まり**

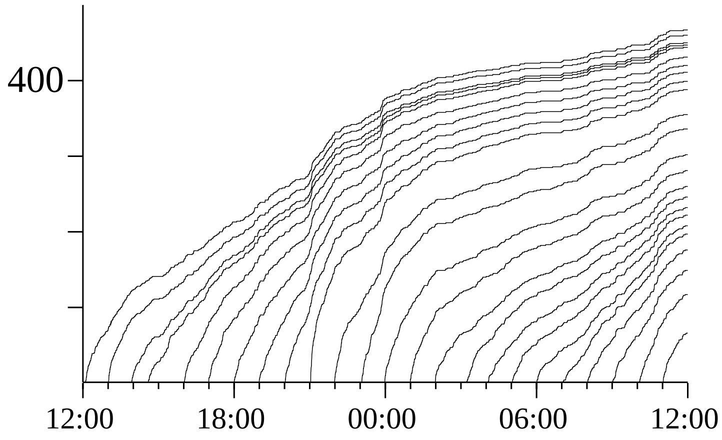
**スレッド（ページ）一覧：**
**書き込んだスレッドが**1**位**
**move-to-front 規則**
**（注：**sage**進行は除く）**

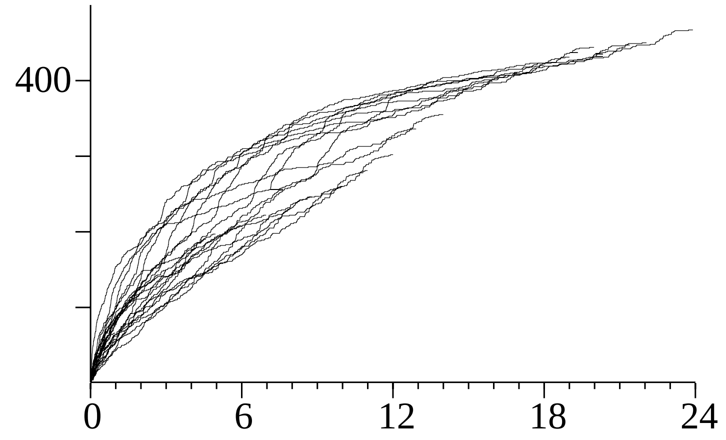



$N$ **は** $10^3$ **程度で小さい（ので，数%の揺らぎがある）が，**
**先頭に跳ぶ規則が明らかなので，踏み込んだ検証が可能**
**スレッド一覧の**1**ページに全スレッドの順位情報→全数調査可能**

# スレッド一覧の順位変化

## 強度の日変化の推定




左図は24個のスレッドの，最後に書き込まれて以降の順位変化

右図は最後に書き込まれた時刻を0に取り直して重ねた図

昼夜差のため，1つの関数$y_C(t)$に重ならない

データ自動収集プログラム竹島佑介君（2008年度東北大M2，現富国生命）

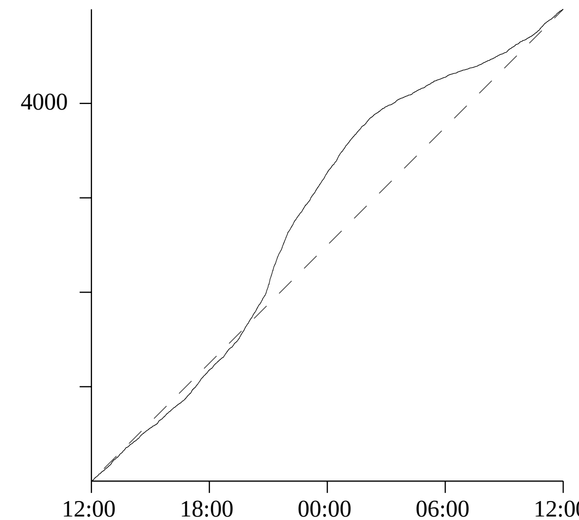
# 強度の共通の時刻依存性と累積書込数

**（再掲）　共通の時刻依存性の仮定：**

$$\rho_i^{(N)}((0,t]) = w_i^{(N)} A(t)$$

**累積総書込数** $S^{(N)}(t) = \sum_{i=1}^{N} \nu_i^{(N)}((0,t])$

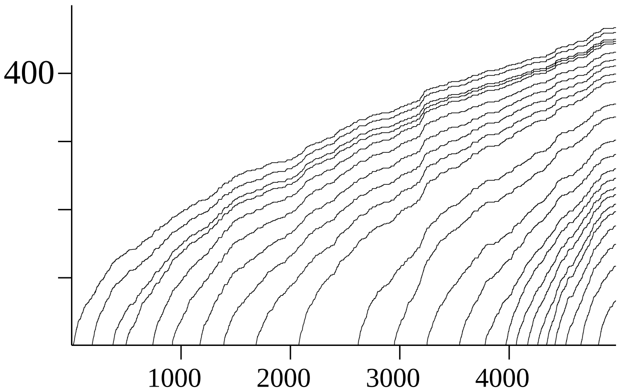
$$\sim A(t)Z(N);\ Z(N) = \sum_{i=1}^{N} w_i^{(N)}$$

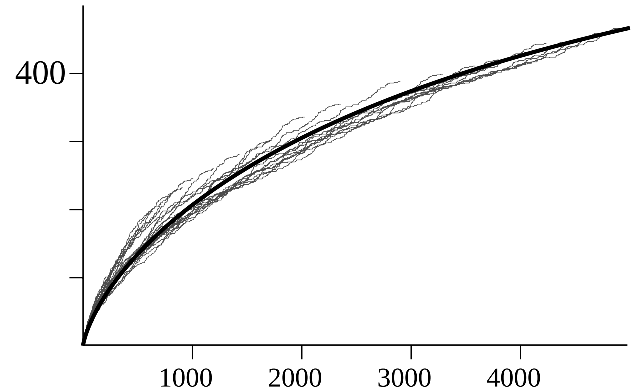



**集計 小林孝長君（**2009 **年度東北大** M2 **，現仙台二高）**

# 共通の時刻依存性の検証とPareto指数の普遍性

（再掲）　$X_i^{(N)}(t) \sim N(1-\int_0^\infty e^{-wS^{(N)}(t)}\lambda(dw));\ \lambda([w,\infty)) = \left(\frac{a}{w}\right)^b,\ w \geqq a$



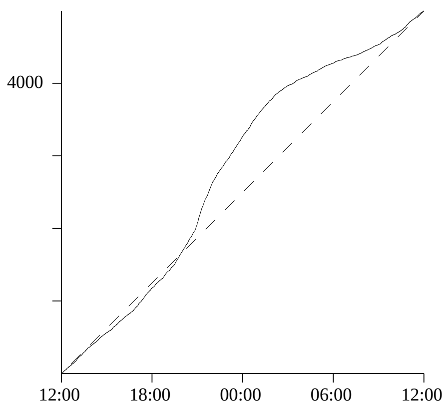

- **共通の時刻依存性は実用的に良さそうである**

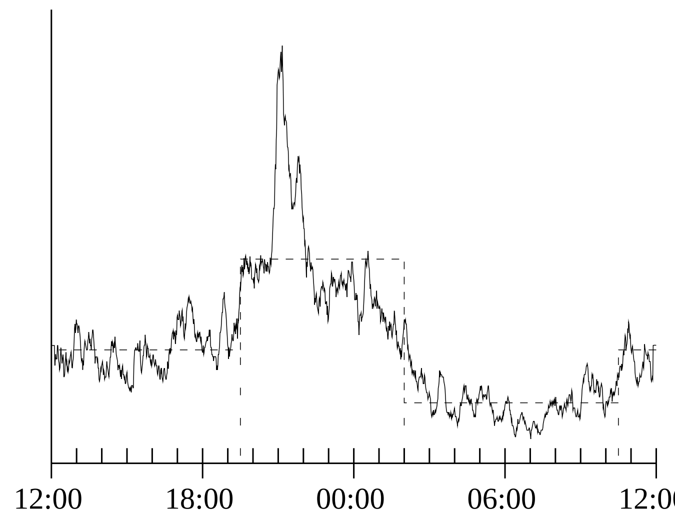
$X_i^{(N)}(A^{-1}(t))$ （$A(t) \sim S^{(N)}(t)$）　→　$b = 0.872 < 1$

- Amazon.co.jp も 2ch.net も共通して $b < 1$

  **本の購入はベストセラーに，書き込みは人気スレッドに集中**

# 昼夜差



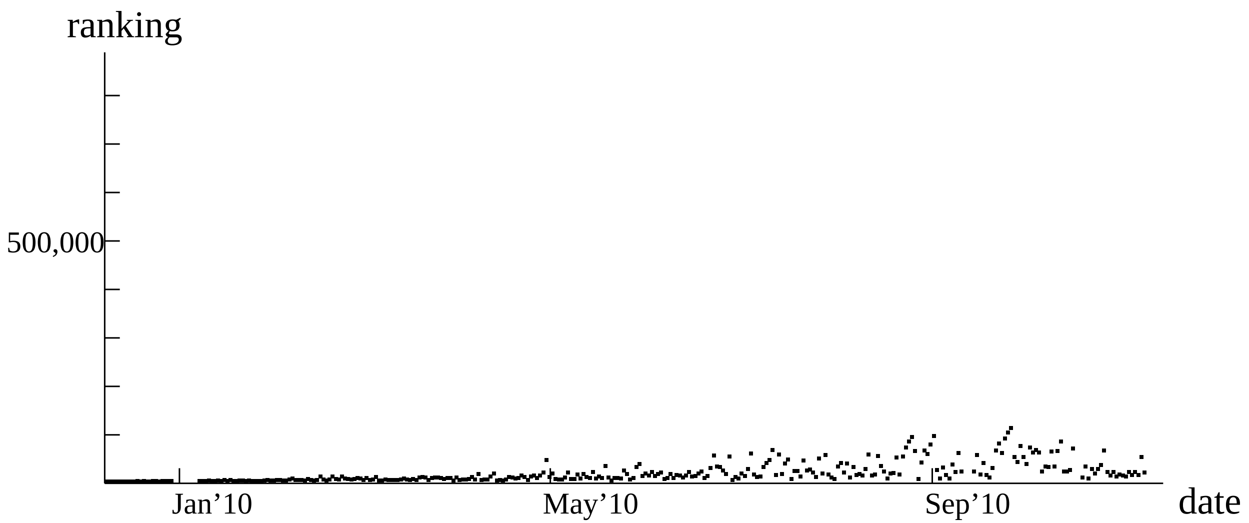

- 20–01時に書込活動活発，03–09時に不活発（２ちゃんねる）

（アマゾン書店と整合）

# ５－１． 普遍性とロングテール

アマゾン書店のアルゴリズムは確率順位付け模型か？（多分違う）
なぜ合うか？ 普遍性の仮説：
めったに売れない本の流行度は最後に売れた順しかない

流体力学的極限：大数の法則 → 大多数の（i.e., 普通の）本
＝めったに売れない本 ＝ロングテール




左半分（刊行間もなく）１万位未満，右半分７万位未満
ロングテールで問題になる順位は，
常識的に順位を考える領域を１位と同一視する縮尺（新しい現象）

## ５－２．　確率的な順位付け

（再掲）　下位の $w_i^{(N)}$ は（データ収集の意味で）測定可能量ではない

データから決めるには $1/w_i^{(N)} \gg 1$ の時間を要する　「五十歩百歩」

特に，流行度を決めるのに時間をかけることは許されない

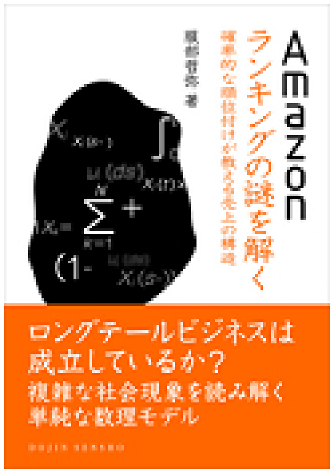
順位付けは確率的な概念である．

それを人は知っているから「試合で決する」という表現がある

ロングテールは順位付けが確率的であることが顕著な領域を扱う

流体力学的極限という一つの帰結を得た．他に何があるか？

# 文献




- 出席　１日あたり１点
- 講義感想 A4 レポート用紙 1/2-1 枚　１点
- レポート問題　１題あたり３点

　◯大数の強法則または Martingale 収束定理の証明

　◯各粒子が $A_i(t) = 1 \pm \delta$ を半周期ずつとり，各時刻で半分ずつのとき，軌道は共通一様強度 $A_i(t) = 1$ のときより高いか低いか不定か

　◯順位と量の関係がべき法則で fit できそうな社会現象を探して，平等性の指数 b を算出．手なら 50 以上，ネットなら 500 以上のデータ

　◯ Amazon 新品を買い占めて飢餓感を煽り割高な中古を売る商売が成立つ条件を整理し，ランキングの動きから察知する方法を考察

- 教科書読了者は，本の書評感想 800 和字－2000 和字（コピペと回答のネット募集不可）　２点

**内容を審査の上，以上を合計して，５点以上で「優」．**

**服部哲弥，「Amazon ランキングの謎を解く」，化学同人，2011.5．**

Google 検索キーワード　　服部哲弥