Stochastic ranking process with time dependent intensities

針谷祐　（東北大・理）
服部久美子　（首都大・数理）
服部哲弥　（慶應大・経済）
永幡幸生　（大阪大・基礎工）
竹島佑介　（富国生命）
小林孝長　（仙台二高）

確率ランキング過程（move-to-front 規則）の，粒子軌道および位置ージャンプ強度結合経験分布についての無限粒子極限の存在定理を，列先頭へのジャンプが非一様な強度を持つ Poisson random measure で記述される場合に拡張する．模型に特徴的なメカニズム（証明の本質）は一様な場合と異ならないが，技術的徹底もあって，一様な場合もより少ない仮定でより強い収束を得た．

ウェブのランキングのデータを確率ランキング模型で説明する場合に，社会の昼夜の活動差を強度の非一様性で説明する．実用上の方法として，強度の時刻依存性が全ての粒子で共通な場合を考える．相対強度の分布を一般化した Pareto 分布として，自動収集によって得た 2ch.net のスレッド一覧の 1 日分のデータを当てはめた結果は，この方法の有効性を示唆する．また，人気スレッドへの書き込みの集中（非ロングテール的），および，深夜に活動が活発で早朝に少ないこと，という，Amazon.co.jp と共通の結果を得た．

1. Poisson random measure で定まる確率ランキング模型の無限粒子極限．

$\mathcal{M}(\mathbb{R}_+)=\{\mathbb{R}_+$上の Radon 測度$\}$ とおく．N を自然数，$\nu_i^{(N)},\ i=1,2,\cdots,N$, を確率空間 $(\mathrm{P},\mathcal{F},\Omega)$ で定義された $\mathbb{R}_+$ 上の Poisson random measures とする．$\nu_i^{(N)}$ の intensity measure を $\rho_i^{(N)}\in\mathcal{M}(\mathbb{R}_+)$ $(\mathrm{E}[\,\nu_i^{(N)}(A)\,]=\rho_i^{(N)}(A))$ とおき，連続である $(\rho_i^{(N)}(\{t\})=0,\ t\geqq 0)$ とする．$x_1^{(N)},x_2^{(N)},\cdots,x_N^{(N)}$ を $1,2,\cdots,N$ の順列とするとき，これを初期値とする非一様な強度を持つ stochastic ranking process $X^{(N)}=(X_1^{(N)},\cdots,X_N^{(N)})$ を，$i=1,2,\cdots,N$ と $t\geqq 0$ に対して

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

で定義する．ここで $\mathbf{1}_A$ は事象 $A\in\mathcal{F}$ の定義関数．確率 1 で，$\nu_i^{(N)}=\sum_{j=1}^{\infty}\delta_{\tau_{i,j}^{(N)}}$ と書ける．ここで，δ_a は a に集中した単位分布，$\tau_{i,j}^{(N)}$ は j について狭義増加する正値確率変数で，(i,j) に関して全て互いに異なる．強度が一様 $(\rho_i^{(N)}((0,t])=w_i^{(N)}t)$ のとき，既報告済のモデル [1,2,3] に一致する．

位置変数をスケールして $[0,1)$ に収める：$Y_i^{(N)}(t)=N^{-1}(X_i^{(N)}(t)-1)$．ジャンプ済み粒子と未ジャンプ粒子の境界位置は $Y_C^{(N)}(t)=\dfrac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\tau_{i,1}^{(N)}\leqq t}$ で表される．

命題([4, Prop. 1])．$t\geqq 0$ とする．分布 λ_t が存在して $\lambda_t^{(N)}=\dfrac{1}{N}\sum_{i=1}^{N}\delta_{\rho_i^{(N)}((0,t])}\to\lambda_t$, weakly as $N\to\infty$, ならば，$Y_C^{(N)}(t)\to y_C(t)=1-\int_0^{\infty}e^{-s}\lambda_t(ds)$, a.e.. ◇

$\mathcal{M}(\mathbb{R}_+)$ はvague位相を入れると，$\mathcal{M}(\mathbb{R}_+)\times[0,1)$ はPoland空間である．$t \geqq 0$ に対して，位置－ジャンプ強度結合経験分布を $\mu_t^{(N)} = \frac{1}{N}\sum_{i=1}^{N} \delta_{(\rho_i^{(N)}, Y_i^{(N)}(t))}$ とおく．

定理([4, Thm. 3])．分布 μ_0 が存在して $\mu_0^{(N)} \to \mu_0$, weakly, および $\Lambda(d\rho) = \mu_0(d\rho \times [0,1))$ とおくとき $\frac{1}{N}\sum_{i=1}^{N} \delta_{\rho_i^{(N)}((s,t])} \to \lambda_{s,t} = \int_{\mathcal{M}(\mathbb{R}_+)} \delta_{\rho((s,t])}\Lambda(d\rho)$, weakly, とする．このとき，各 $t>0$ に対して，確率1で，$\mu_t^{(N)}$ は分布 μ_t に弱収束する．μ_t は非ランダムで，以下で与えられる：

$$U(d\rho, y, t) = \mu_t(d\rho \times [y,1)) = \begin{cases} e^{-\rho((t-t_0(y,t),t])}\,\Lambda(d\rho), & 0 \leqq y \leqq y_C(t), \\ e^{-\rho((0,t])}\,U(d\rho, \hat{y}(y,t), 0), & y_C(t) \leqq y < 1. \end{cases}$$

ここで，$t_0(y,t)$ は $y_A(t_0,t) = 1 - \int_{\mathcal{M}(\mathbb{R}_+)} e^{-\rho((t-t_0,t])}\,\Lambda(d\rho)$ の t_0 に関する逆関数, $\hat{y}(y,t)$ は $y_B(y,t) = 1 - \int_{\mathcal{M}(\mathbb{R}_+)} e^{-\rho((0,t])}\,\mu_0(d\rho \times [y,1))$ の y に関する逆関数． ◇

以上において，$\{\rho_i^{(N)}\}$ などの同程度連続性に関する仮定をおけば，確率過程としての概収束も得る．

2．共通の時刻依存性を持つ intensity measure とウェブ上の活動の昼夜差．

強度が一様の場合の軌道 y_C は，ウェブのランキングで毎日定時の観測や１日に満たない観測について実際に観測された [2,3]．非一様な場合に定理を拡張したことによって，社会の昼夜の活動差が無視できないデータを解析することができる．もっとも単純な方法として，密度 $\tilde{a} \in L^1_{loc}(\mathbb{R}_+)$ に基づく共通の時刻依存性 $\rho_i^{(N)}((0,t]) = w_i^{(N)} \int_0^t \tilde{a}(u)\,du =: w_i^{(N)} A(t)$ と $\frac{1}{N}\sum_{i=1}^{N}\delta_{w_i^{(N)}} \to \lambda$ を仮定する．このとき $y_C(t) = 1 - \int_{\mathbb{R}_+} e^{-w\,A(t)}\,\lambda(dw)$ となる．特に，$\tilde{a}$ が１日周期ならば毎日定時のデータは一様な場合の公式に帰着するので，以前の解析 [3] が正当化される．さらに，多数の軌道のデータがあるとき，$\tilde{a}$ を分離して λ を統計的に解析できる．

定理([4, Thm. 15])．$\tilde{a}$ は正で $Z(N) = \sum_{i=1}^{N} w_i^{(N)} \to \infty$ のとき $Y_C^{(N)}(s^{(N)}(Z(N)\,t)) \to y_C(A^{-1}(t))$ in prob. ここで，$s^{(N)}$ は $S^{(N)}(t) = \sum_{i=1}^{N} \nu_i^{(N)}((0,t])$ の逆関数． ◇

一般化Zipfの法則 $w_i^{(N)} = a(N/i)^{1/b}$（λ がPareto分布）に当てはめて，2ch.net のあるスレッド一覧の１日の全数データ $(N = 697)$ から $b = 0.872 \pm 0.002$ を得た．

文献（`http://web.econ.keio.ac.jp/staff/hattori/liamazn.htm`）

[1] K. Hattori, T. Hattori, Stochastic Processes and their Applications **119** (2009) 966–979.
[2] K. Hattori, T. Hattori, Funkcialaj Ekvacioj **52** (2009) 301–319.
[3] K. Hattori, T. Hattori, preprint (2008).
[4] Y. Hariya, K. Hattori, T. Hattori, Y. Nagahata, Y. Takeshima, T. Kobayashi, *Stochastic ranking process with time dependent intensities,* preprint (2010).