

Convergence du classement **Elo**, du microscopique au mésoscopique

Stéphane Junca

JAD, Inria & CNRS, Université Côte d'Azur

Porquerolles 28 Juin 2021
Nice-Toulon-Marseille

joint work: Pierre-Emmanuel Jabin, Penn State University,
Adrien Krifa, Florian Spinelli, Master IM, Université Côte d'Azur

1 The Elo: $R_i^{n+1} - R_i^n = K(S_{ij}^{n+1} - b(R_i^n - R_j^n))$

2 Markov chain: N players

3 ODE: $KL N \ll 1$, $L = \sup b'$

4 PDE: Continuity equation: $N \rightarrow \infty$

Arpad Elo's rating for 2 players

Player number i meets player number j

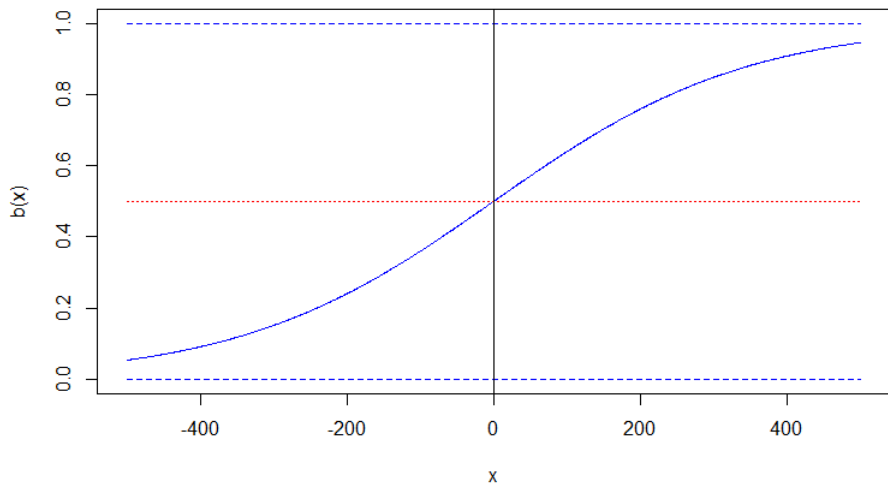
Update the ratings:

$$R_i^{n+1} - R_i^n = K(S_{ij}^{n+1} - b(R_i^n - R_j^n))$$

$$R_j^{n+1} - R_j^n = K(S_{ji}^{n+1} - b(R_j^n - R_i^n))$$

- R_i^n : rating of player number i at time $n\Delta t$
- $0 < K$ 'small'
- **Score** for player i after interaction with the player j
 $S_{ij}^{n+1} \in \{0, 1\}$ or $\{0, 1/2, 1\}$ in practice
or the centered case $S_{ij}^{n+1} \in [-1, 1]$ to see symmetries
- $b(\cdot)$ **damping**: $b(\cdot) = 0$, random walk, divergence of ratings
 $b(\cdot)$ bonus-malus function, increasing and odd function
 $b(\cdot)$ as the **expected score**: $\mathbb{E}(S_{ij}^{n+1}) = b(R_i^n - R_j^n)$

function b



$$R_i^{n+1} - R_j^n = K(S_{ij}^{n+1} - b(R_i^n - R_j^n)).$$

- 1 invariance by translation :

$$R \leftrightarrow R + \text{constant}$$

Only the relative rating is relevant

- 2 "mass" conservation:

$$R_i^n + R_j^n = R_i^0 + R_j^0$$

- $S_{ij}^n = -S_{ij}^{n+1}$
- The function $b(\cdot)$ is **odd** and monotone

Elo assumptions and convergence

$$R_i^{n+1} - R_i^n = K(S_{ij}^{n+1} - b(R_i^n - R_j^n)).$$

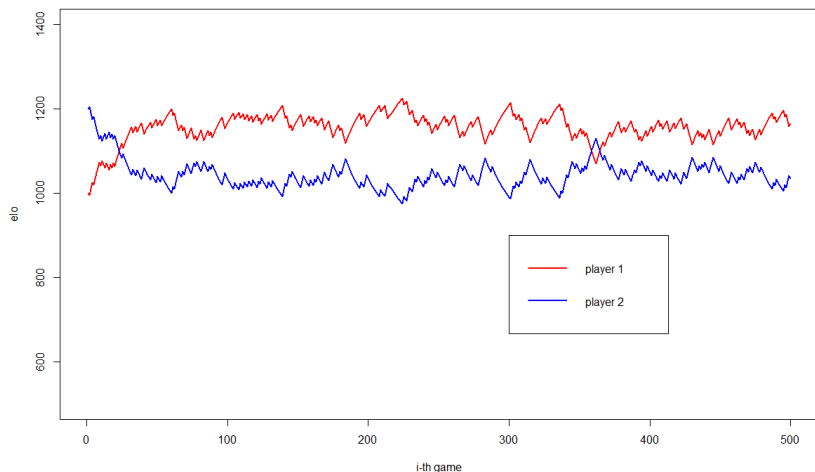
- 1 ϱ_i : "true" rating **fixed**: no improvement, no regression, **unknown**
- 2 S_{ij}^{n+1} : score for player i after interaction with the player j

$$\mathbb{E}(S_{ij}^{n+1}) = b(\varrho_i - \varrho_j)$$

Independent scores w.r.t. $n, i, j \implies$ a Markov chain

$$\lim_{n \rightarrow +\infty} \mathbb{E}(R_i^n) = \varrho_i \quad ??? \quad (\text{if } R_i^0 + R_j^0 = \varrho_i + \varrho_j)$$

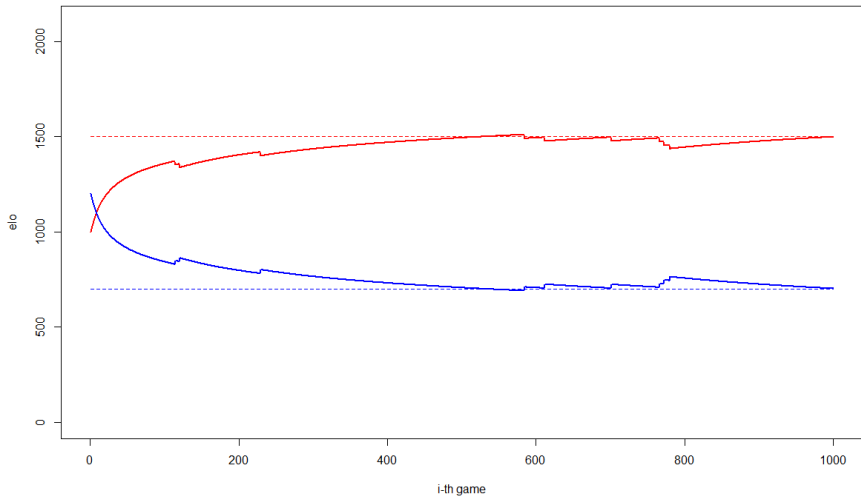
Elo Simulation, 500 games, 2 players



$p_{12} = 60\% = b(72) R_1^0 = 1000 < R_2^0 = 1200, \rho_1 = 1136 > \rho_2 = 1064.$

Master project 2021: Adrien Krifa, Florian Spinelli

ELO Simulation, with $p = b(\rho_1 - \rho_2)$



$$\rho_1 - \rho_2 = 1000,$$

$$p_{12} = b(\rho_1 - \rho_2) \approx 99\%$$

First critics on hypothesis $p_{ij} = b(\varrho_i - \varrho_j)$

Probabilistic interpretations:

- $S_{ij} \sim \mathcal{B}(p_{ij})$: $S_{ij}^{n+1} = 0$ or 1 , Bernoulli law
- p_{ij} : fixed probability that player i beats player j

• $\varrho_1, \dots, \varrho_N$, $(N - 1)$ free parameters

$(p_{ij})_{1 \leq i < j \leq N}$, $\frac{N}{2}(N - 1)$ free parameters

$N = 2$ players, OK

$N \geq 3$ overdetermination of p_{ij}

$$R_i^{n+1} = R_i^n + K(S_{ij}^{n+1} - b(R_i^n - R_j^n)).$$







- $N = 2$,
 - $R_i \sim \mathcal{N}(\rho_i, \sigma^2)$, $\sigma = 200$,
 - b ("centered") normal distribution $\mathcal{N}(0, 2\sigma^2)$
 $b(\cdot) = \text{Erf}(\cdot/(\sqrt{2}\sigma))$ or $2 \text{Erf}(\cdot/(\sqrt{2}\sigma)) - 1$
- $N > 2$, **numerics** convinced the American Chess Federation
- $N \geq 3$, b to be replaced by a **logistic** (centered) function

$$\text{Transitivity} \implies b(r) = \frac{1}{1 + \exp(-c r)} \quad (-1/2),$$
$$c = \frac{\ln(10)}{400} \simeq 0.6 \cdot 10^{-3}$$

- Strength: $F_{ij} = \frac{N_{ij}}{N_{ji}}$ with N_{ij} the number of win for i against j .
- Transitivity $F_{ij} F_{jk} = F_{ik} \implies \ln(F_{ij}) = c(\varrho_i - \varrho_j)$
- Expected score: $p_{ij} = \frac{N_{ij}}{N_{ij} + N_{ji}} = \frac{F_{ij}}{1 + F_{ij}} = b(\varrho_i - \varrho_j)$

A success story

- American Chess federations, FIDE $b(\cdot)$ logistic
- Football Univ. teams, Basket teams, USA, Australia
- Facemash 2003 before Facebook
- FIDE Chess: 3 400 000 players
- Online Games
 - LiChess: 552 176
 - World of Warcraft: 5 000 000
 - Chess.com: 50 000 000
 - Rocket League: 53 000 000
 - League of Legends : 115 000 000
 - Fornite: 350 000 000
- Data, informatics

-  Bradley, R. A.; Terry, M. E. The method of paired comparisons. *Biometrika* 39 (1952), 324-345.
-  Elo, A.: The rating of chessplayers, past and present. 1978, 206 p.
-  Glickman, Mark. Statistics, Harvard Phd, 1993 → Glicko
-  **Avdeev**, V. A.: *Stationary distribution of the player rating in the Elo model with one adversary*. *Diskret. Mat.* 26 (2014), no. 4, 3-14
-  Jabin, P.-E.; Junca, S.: *A continuous model for ratings*. *SIAM J. Appl. Math.* 75 (2015), no. 2, 420–442.
-  **Aldous**, D.: *Elo ratings and the sports model: a neglected topic in applied probability?* *Statist. Sci.* 32 (2017), no. 4, 616–629.

$$R_i^{n+1} - R_i^n = K \sum_j W(R_i^n - R_j^n) (S_{ij}^{n+1} - b(R_i^n - R_j^n)).$$

$0 \leq W(\cdot)$ even function

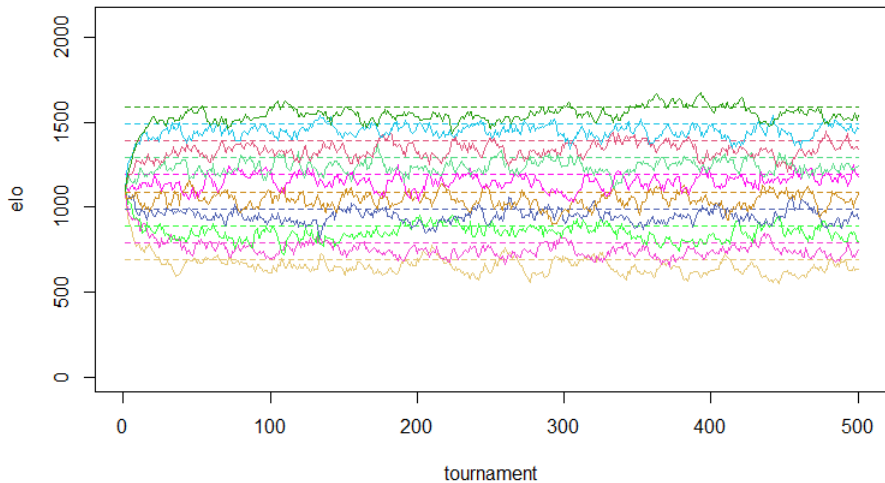
- 1 $W \equiv 1$: all interact with all. Round-robin tournaments.
- 2 Interactions between players having almost the same rating,

$$W(\Delta r) = \begin{cases} 1 & \text{if } |\Delta r| < 100 \\ 0 & \text{else .} \end{cases}$$

- 3 $W \sim \mathcal{B}(p)$, $0 < p < 1$.

Markov chain (independent)

ELO evolution and rho convergence for N players



Linear bonus $b(r) = Lr$

In chess practice: $L = b'(0) \sim 1,410^{-3}$ for the logistic law
valid approximation for $\Delta R < 400$
Used to compute the Elo performance

$$R_i^{n+1} - R_i^n = K \sum_j (S_{ij}^{n+1} - b(R_i^n - R_j^n))$$
$$R^{n+1} = AR^n + K S^{n+1} \mathbf{1}$$

Ratings Vector $R^n = (R_i^n)_{i \leq N}$
Scores matrix $S^{n+1} = (S_{ij}^{n+1})$
 $\mathbf{1}^\top = (1, \dots, 1)$

Mas conservation: $\mathbf{1}^\top R^n = \mathbf{1}^\top R^0$

Invariance by translation: $R^n \leftrightarrow R^n + c \mathbf{1}$

$$R^{n+1} = A^n R^0 + K \sum_{k \leq n} A^k S^{n+1-k} \mathbf{1}$$

Contractance on $\mathbb{1}^\perp$ if $KLN < 2$ then $\gamma = 1 - KLN \in]-1, 1[$ and

Theorem (Geometric convergence in law, J.)

$$\begin{aligned}R^n &\rightharpoonup \frac{\mathbb{1}^\top R^0}{N} \mathbb{1} + R^\infty \\ \mathbb{1}^\perp R^\infty &= K \sum_{n=0}^{\infty} \gamma^n S^{n+1} \mathbb{1} \\ \mathbb{E}(R^\infty) &= \frac{1}{LN} \sum_{j \neq i} \mathbb{E}(S \mathbb{1})\end{aligned}$$

Arpad Elo assumption is confirmed in the linear case:

$$\varrho = (\varrho_1, \dots, \varrho_N)^\top, \quad \mathbb{1}^\top \varrho = \mathbb{1}^\top R^0$$

If $p_{ij} = b(\varrho_i - \varrho_j)$ then $\mathbb{E}(R_i^n) \rightarrow \varrho_i$ for all i .

Theorem (Nonlinear contractive case, J .)

If $K(N-1) \sup \mathbf{b}' < 1$ and $0 < \inf \mathbf{b}'$
then, contractance: $\gamma = \max(|1 - KN \sup \mathbf{b}'|, |1 - KN \inf \mathbf{b}'|) < 1$,
and convergence in law towards a unique invariant measure.

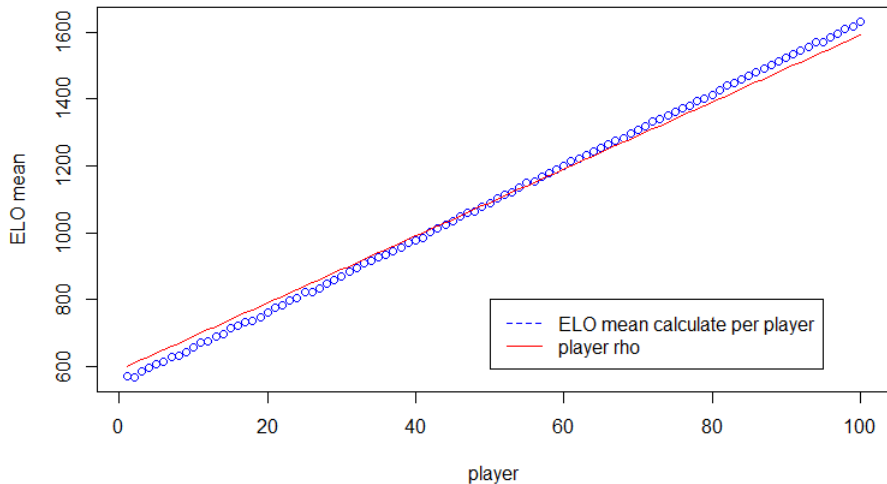
Drawbacks:

- 1 For the **logistic** or a distribution function, $\inf \mathbf{b}' = 0$.
- 2 $\inf \mathbf{b}' > 0 \Rightarrow \lim_{r \rightarrow \pm\infty} \mathbf{b}(r) = \pm\infty$.
- 3 **Decay of too large ratings after a win** against a too low rating
If $\mathbf{b}(R_i^n - R_j^n) \gg 1$ then $R_i^{n+1} - R_j^n \sim -\gamma K < 0$.

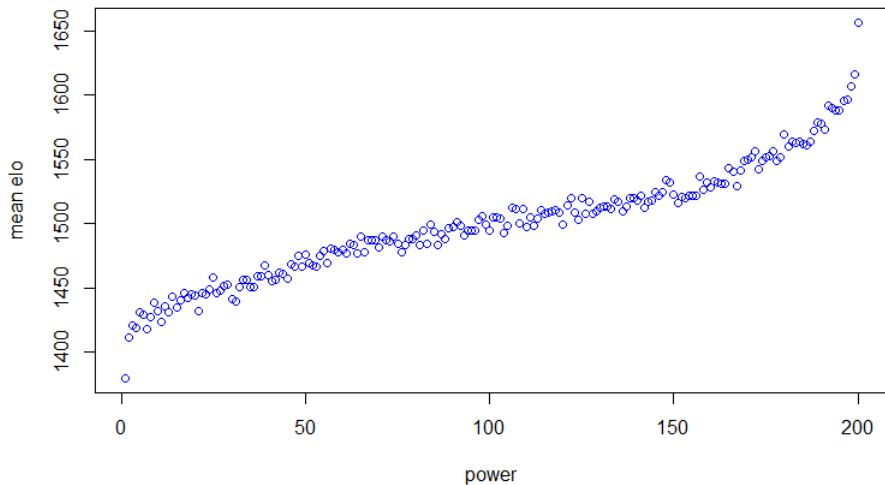
Advantages:

- 1 Linear case
- 2 Geometric convergence, \mathbf{b} as a damping
- 3 Uniformly bounded ratings, attractive zone
- 4 Drawbacks 3. never occurs (except for the best and the weakest)

ELO mean of 100 players in RRT with update after each tournaments



Means of ELO for P free sorted by theoretical power



$$\frac{R_i^{n+1} - R_i^n}{\Delta t} = \sum_j \frac{K}{\Delta t} W_{ij}^n (R_i^n - R_j^n) (S_{ij}^{n+1} - b(R_i^n - R_j^n)).$$

After normalization, score replaced by its mean, chain by process,

$$w(r) = \frac{K}{\Delta t} \mathbb{E}(W_{ij}^n)$$

$$\frac{dR_i(t)}{dt} = \sum_j w(R_i(t) - R_j(t)) (b(\varrho_i - \varrho_j) - b(R_i(t) - R_j(t)))$$

- Many interactions $\gg 1$
- $\Delta t \ll 1$
- $K \ll 1$

Empiric measure: $f(t, r, \varrho) = \frac{1}{N} \sum_i \delta(r - R_i(t)) \delta(\varrho - \varrho_i)$

$$\begin{cases} \frac{\partial f}{\partial t} + \frac{\partial}{\partial r} (a[f] f) = 0, \\ f(0, r, \varrho) = f_0(r, \varrho), \end{cases}$$

mean field velocity:

$$a[f](t, r, \varrho) = \iint_{\mathbb{R}^2} (w(r - r') (b(\varrho - \varrho') - b(r - r'))) f(t, r', \varrho') d\varrho' dr'.$$

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial r} (a[f] f) = 0, \quad f(0, r, \varrho) = f_0(r, \varrho),$$

- $\iint f(t, r, \varrho) dr d\varrho = 1$
- Marginals: $f(t, r, -) = \int f(t, r, \varrho) d\varrho$,
 $f(t, -, \varrho) = \int f(t, r, \varrho) dr = f_0(-, \varrho)$
- means: $\int r f(t, r, -) dr = m_1 + \int \varrho f_0(-, \varrho) d\varrho$,
"mass conservation" & invariance by translation
- **Entropies** \downarrow : $E_k(t) = \iint |r - \varrho|^k f(t, r, \varrho) dr d\varrho$

Theorem (convergence in $W^{-1,1}(\mathbb{R}_r \times \mathbb{R}_\varrho)$, Jabin, J.)

Initial density compactly supported

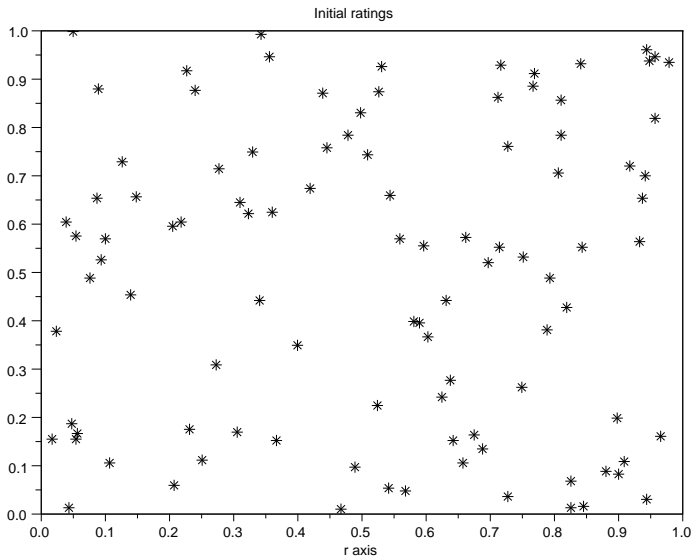
$$f_0 \in M_{\mathbf{c}}^1(\mathbb{R}^2),$$

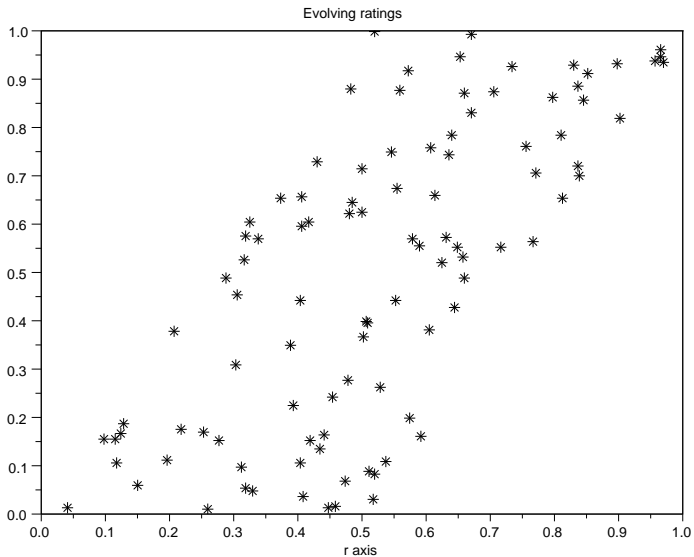
$$\iint f_0(r, \varrho) dr d\varrho = 1$$

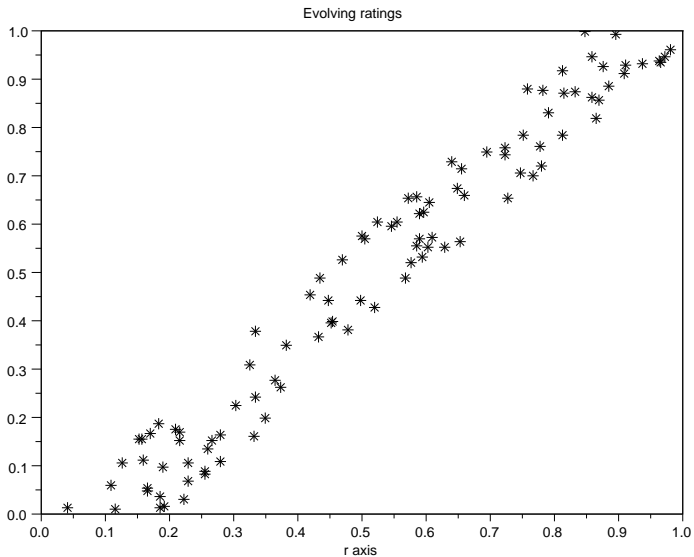
$$\inf_{\mathbb{R}} w > 0$$

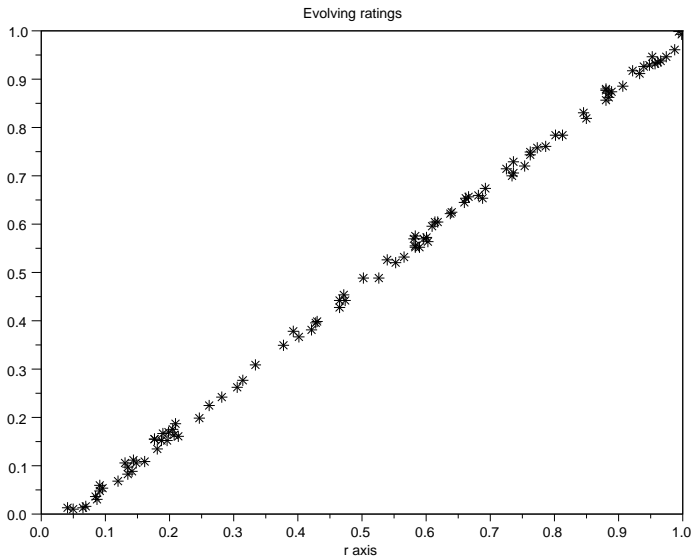
$$\implies \left\| f(t, \cdot, \cdot) - f^0(\cdot, \varrho) \delta(r - \varrho - m_1) \right\|_{W^{-1,1}(\mathbb{R}^2)} \leq C e^{-\lambda t}.$$

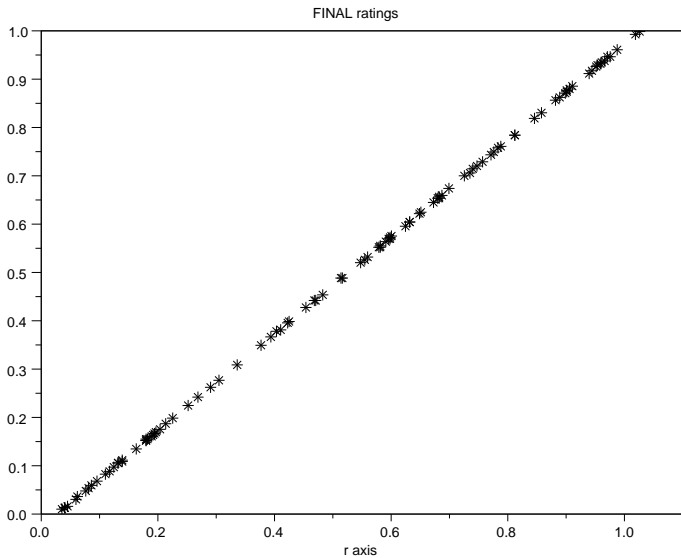
Convergence in all Wasserstein distances $1 \leq p < +\infty$











All against all: Sketch of the proof

- Entropies: $\varphi'' > 0$,

$$E_\varphi(t) = \iint \varphi(r - \varrho) f(t, r, \varrho) dr d\varrho \quad \downarrow$$

- Uniform compact support
- Coercive energy:

$$E_2(t) = \iint |r - \varrho|^2 f(t, r, \varrho) dr d\varrho$$

- Dissipation of the energy: $D_\varphi(t) := \frac{dE_\varphi}{dt}$

Dissipation of energy: $D_\varphi(t) = -D(f)$

$$D(f) = \int (r - r' - \varrho + \varrho') w(r - r') (b(\varrho - \varrho') - b(r - r')) f' f dr' dr d\varrho' d\varrho.$$

$$(r - r' - \varrho + \varrho') w(r - r') (b(\varrho - \varrho') - b(r - r')) \geq \beta \underline{w} |r - r' - \varrho + \varrho'|^2.$$

$$D(f) \geq \beta \underline{w} \int |r - r' - \varrho + \varrho'|^2 f' f dr' dr d\varrho' d\varrho.$$

$$\int (r - \varrho)(r' - \varrho') f' f dr' dr d\varrho' d\varrho = 0.$$

$$\frac{d}{dt} E(t) \leq -2 \underline{w} \beta \int |r - \varrho|^2 f = -2 \omega \beta E(t),$$

Local interactions

The energy is no longer coercive

Discrete case, ODE: $f = \sum_{i=1}^N \alpha_i \delta(t - R_i(t)) \delta(\varrho - \varrho_i)$,

Theorem (convergence in $W^{-1,1}(\mathbb{R}_r \times \mathbb{R}_\varrho)$, Jabin, J.)

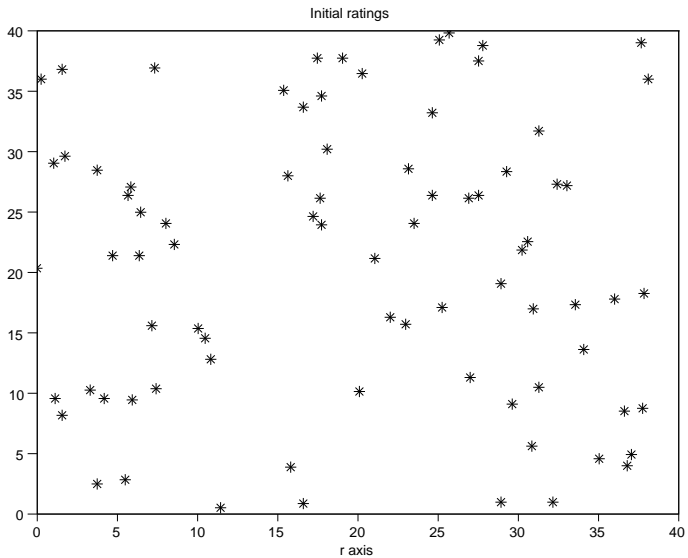
If $0 = w$ outside $[-1, 1]$, $0 < w$ on $]-1, 1[$ and $w \in C_c^0(\mathbb{R}, \mathbb{R})$, then $f(t, r, \varrho)$ converges for large time. Moreover, there exist c_1, \dots, c_N s.t.

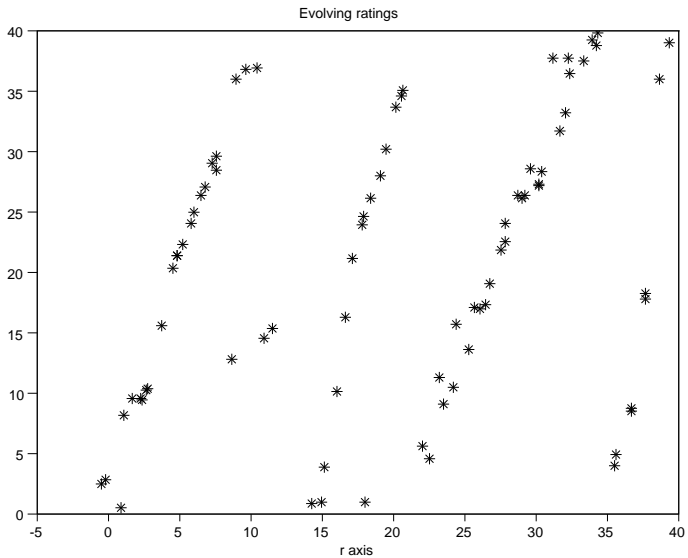
$f(t, \dots) \longrightarrow \sum_{i=1}^N \alpha_i \delta(r - \varrho_i - c_i) \delta(\varrho - \varrho_i)$ in weak- $*$ $M^1(\mathbb{R}^2)$, as $t \rightarrow +\infty$,

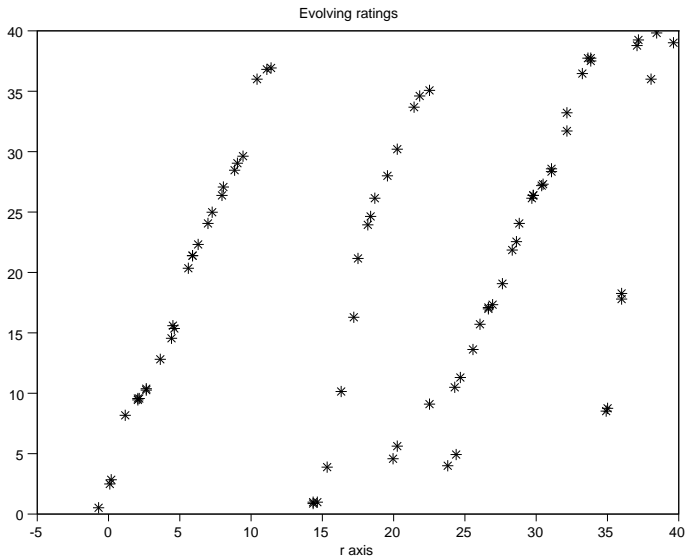
where either $c_i = c_j$ or $|\varrho_i + c_i - \varrho_j - c_j| \geq 1$.

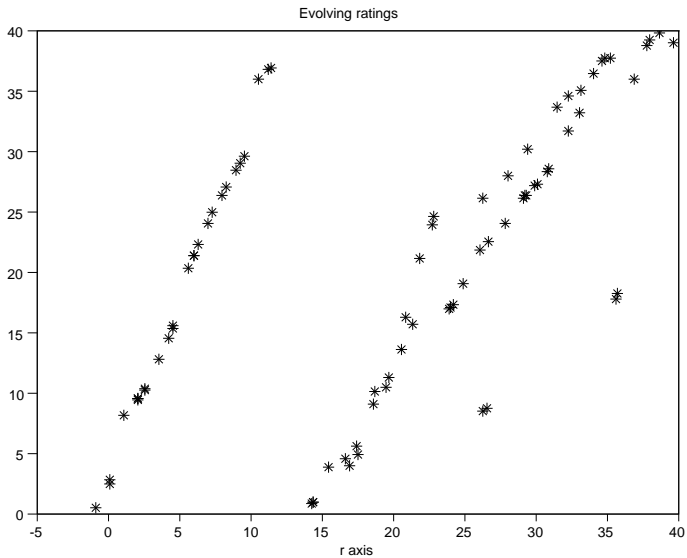
i.e. $R_i(t) \rightarrow \varrho_i + c_i$

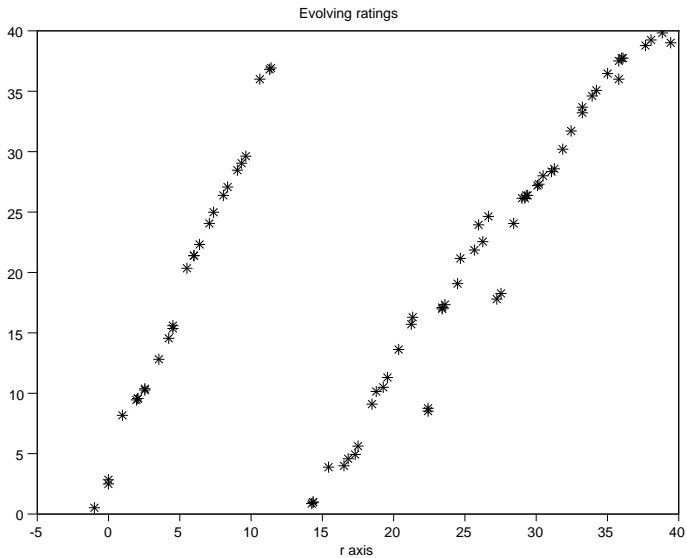
& non-interacting groups for different c_i



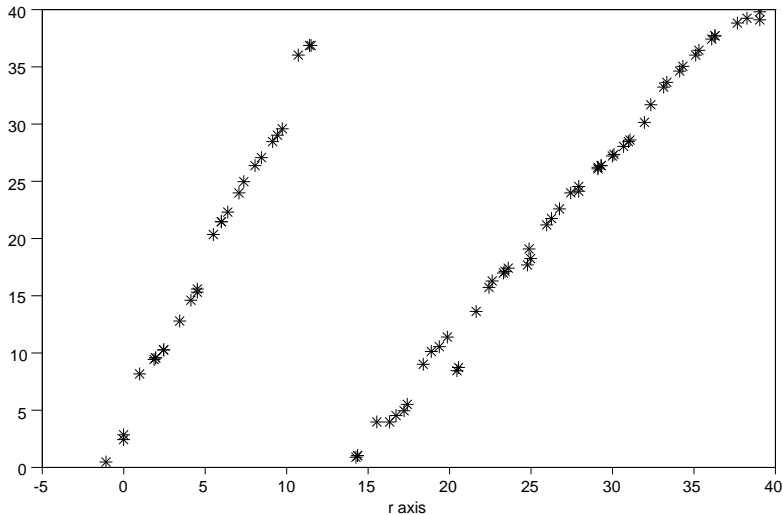








Evolving ratings



Discrete case: sketch of the proof

ODE: N entropies converge ... finite set of limits ...

Kinetic equation: $e_i(t) = R_i(t) - \varrho_i$

- For all $\phi \in C^0(\mathbb{R}, \mathbb{R})$, the function

$$V[\phi(t)] = \sum_{i=1}^N \alpha_i \phi(e_i(t))$$

has a limit as $t \rightarrow \infty$.

- $e_i(t)$ converges for all i
- Identification of the possible limits with the dissipation of the energy: $D_2(+\infty) = 0$.

Convergence for a degenerate kinetic equation

More degenerate than usual: $w(x) \sim (x - x_i)^p$
but with more symmetries

- All entropies converges: $t \rightarrow +\infty$

$$g(t, r) = \int_{\mathbb{R}} f(t, r + \varrho, \varrho) d\varrho \rightarrow g_{\infty}(r)$$

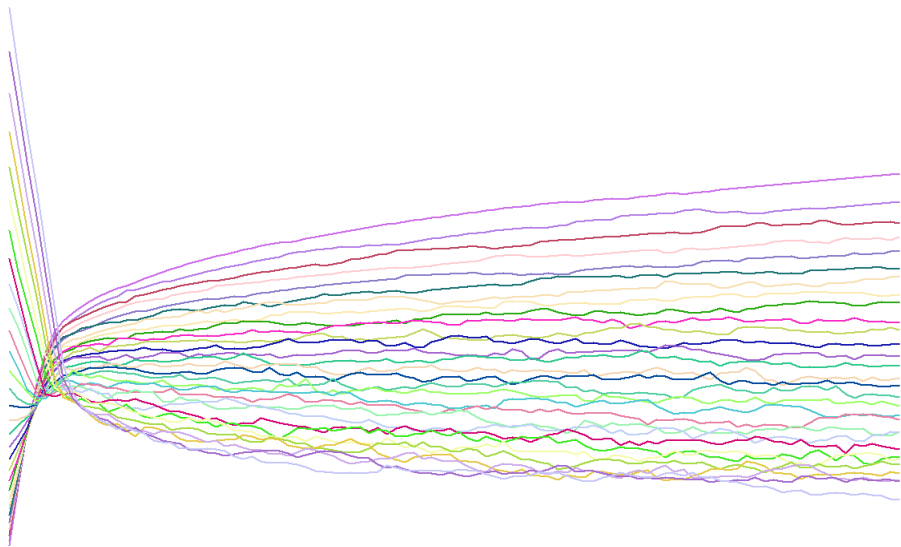
- Identification of the limit with zero dissipation:

$$D_{\varphi} \equiv 0$$

- very weak stability of the possible limit :

$$w \in C^2$$

- 1 On the **Markov chain**:
 - rate of convergence. Are 30 matches enough?
Work in progress with P.-E. Jabin
 - other pairings
 - Vectorial "true" strength: $\mathbb{E}(S_{ij}^{n+1}) = b(\varrho_i^1 - \varrho_j^1, \dots, \varrho_i^d - \varrho_j^d,)$
- 2 On the **kinetic Model**,
 - Identifications of limits w.r.t. initial data for local interactions.
 - Hydrodynamic model
- 3 Other
 - 1 Elo initialisation
 - FIDE uses a linear estimator
 - a new nonlinear estimator
 - 2 Elo Inflation and increasing number of players
 - 3 No fixed strengths: improvement, regression



Thank you
for
your attention !