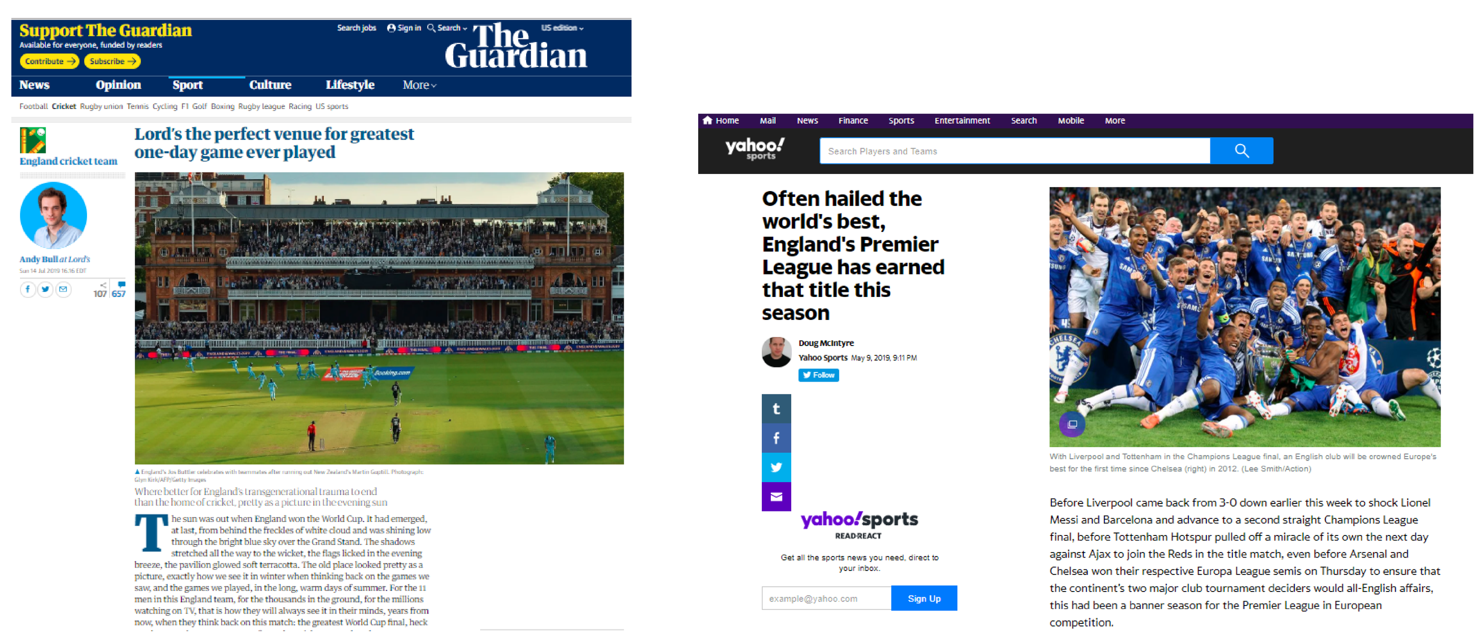
Estimation of Skill Distribution from a Tournament

Ali Jadbabaie, Anuran Makur, and Devavrat Shah

Laboratory for Information & Decision Systems, Massachusetts Institute of Technology

Email: jadbabai@mit.edu, a_makur@mit.edu, devavrat@mit.edu

Motivation



- Can we formalize such claims about sports using statistical analysis?
- Can we measure the overall level of skill in a game based on win-loss data from tournaments?

Formal Setup and Goal

- Unknown **PDF of skill levels P_α** on interval $[\delta, 1]$ for $\delta > 0$, which is bounded and belongs to η-Hölder class.
- Teams $\{1, \ldots, n\}$ play tournament with unknown i.i.d. skill levels $\alpha_1, \ldots, \alpha_n \sim P_\alpha$.
- For $i \neq j$, with probability $p \in (0, 1]$, observe k independent pairwise games where $Z_m(i, j) = \mathbb{I}\{j \text{ beats } i \text{ in } m\text{th game}\}$.
- **Bradley-Terry-Luce (BTL)** or **multinomial logit** model [1]:

$$\mathbb{P}(Z_m(i, j) = 1 \,|\, \alpha_1, \ldots, \alpha_n) = \frac{\alpha_j}{\alpha_i + \alpha_j}.$$

- **Goal:** Learn P_α from observation matrix $Z \in [0, 1]^{n \times n}$ with

$$Z(i, j) = \begin{cases} \dfrac{1}{k} \displaystyle\sum_{m=1}^{k} Z_m(i, j)\,, & \text{if games observed between } i \neq j\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

- **Overall skill score:** Negative differential entropy of P_α

$$-h(P_\alpha) = \int_\delta^1 P_\alpha(x) \log(P_\alpha(x))\,\mathrm{d}x$$

measures the variation of skill levels of teams in a tournament.

- **Intuition:** Concentrated P_α has high score and outcomes of games are unpredictable; Balanced P_α has low score and there is more variation of skill levels.

Estimation Algorithm

Estimation of P_α from Z

Input: Observation matrix Z

Output: Estimator $\widehat{\mathcal{P}}^*$ of P_α

Step 1: **Skill parameter estimation using rank centrality algorithm** [2]

1. Construct stochastic matrix $S \in \mathbb{R}^{n \times n}$ with $S(i, j) = \frac{Z(i,j)}{2np}$ for $i \neq j$, whose rows sum to 1
2. Compute leading left eigenvector $\hat{\pi}_*$ of S such that $\hat{\pi}_* = \hat{\pi}_* S$
3. Compute skill level estimates $\hat{\alpha}_i = \frac{\hat{\pi}_*(i)}{\|\hat{\pi}_*\|_\infty}$ for $i = 1, \ldots, n$

Step 2: **Kernel density estimation using Parzen-Rosenblatt method**

4. Compute bandwidth $h = \Theta\big(\log(n)^{\frac{1}{2\eta+2}} n^{-\frac{1}{2\eta+2}}\big)$
5. Construct $\widehat{\mathcal{P}}^*$ using appropriate, fixed kernel $K : [-1, 1] \to \mathbb{R}$

$$\widehat{\mathcal{P}}^*(x) \triangleq \frac{1}{nh} \sum_{i=1}^{n} K\!\left(\frac{\hat{\alpha}_i - x}{h}\right)$$

6. **Return** $\widehat{\mathcal{P}}^*$

Theoretical Results

Theorem (Mean Squared Error Upper Bound)

If $p = \Omega(\log(n)/(\delta^5 n))$ and $\lim_{n\to\infty} \delta^{-1}(npk)^{-1/2} \log(n)^{1/2} = 0$, then for all P_α,

$$\mathbb{E}\!\left[\int_{\mathbb{R}} \big(\widehat{\mathcal{P}}^*(x) - P_\alpha(x)\big)^2 \mathrm{d}x\right] = O\!\left(\max\left\{\left(\frac{\log(n)}{\delta^2 pkn}\right)^{\frac{\eta}{\eta+1}}, \left(\frac{\log(n)}{n}\right)^{\frac{\eta}{\eta+1}}\right\}\right).$$

- **Summary of all minimax estimation results:**

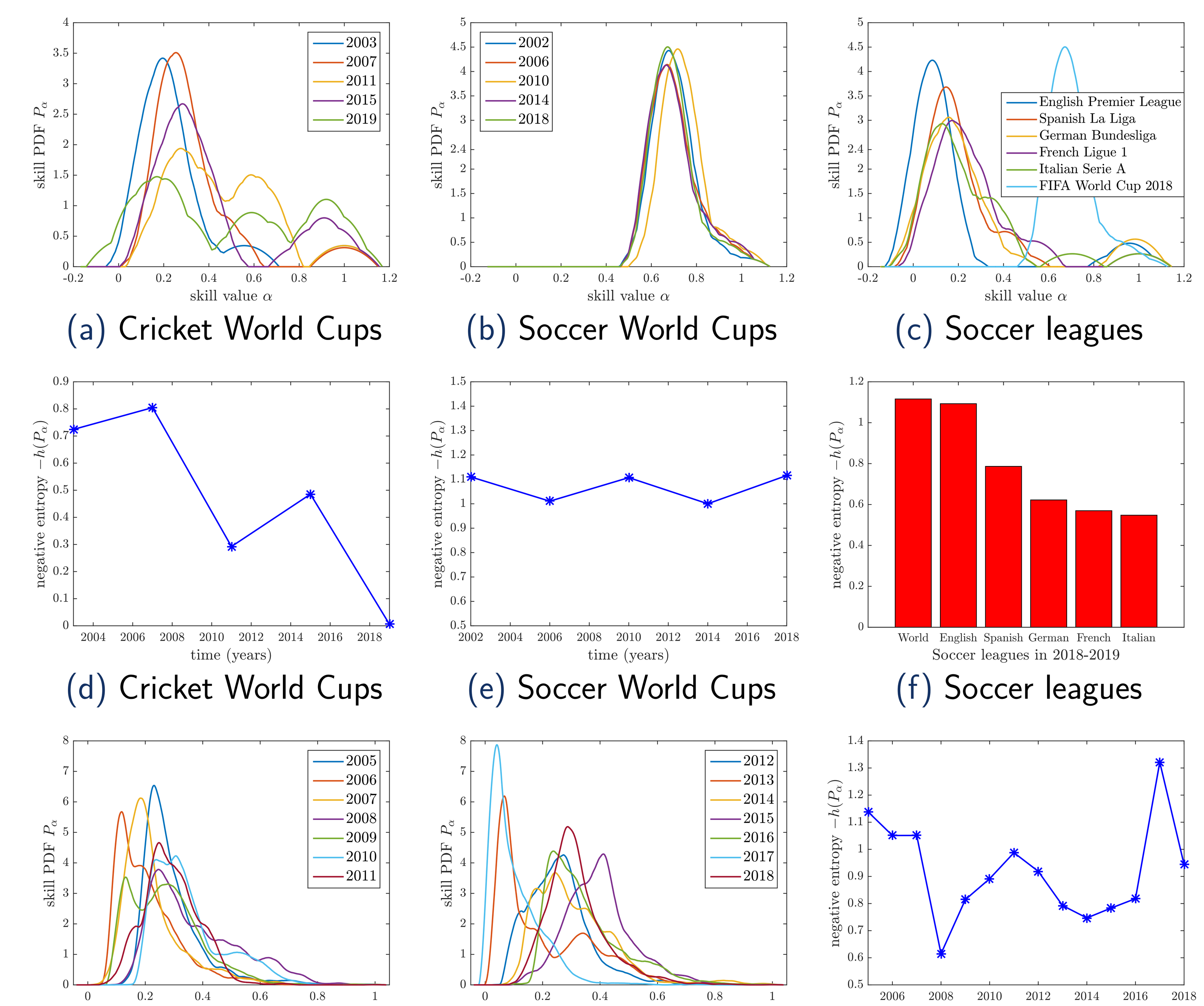
Estimation prob.	Loss func.	Upp. bound	Low. bound
Smooth skill PDF	MSE	$\tilde{O}(n^{-1+\varepsilon})$	$\Omega(n^{-1})$
BTL skill levels	ℓ^∞-norm	$\tilde{O}(n^{-1/2})$	$\tilde{\Omega}(n^{-1/2})$
BTL skill levels	ℓ^1-norm	$O(n^{-1/2})$	$\tilde{\Omega}(n^{-1/2})$

- Note: Our results are in red; other results are known in the literature. Notation $\tilde{O}$ and $\tilde{\Omega}$ hide polylog(n) terms, and $\varepsilon > 0$ is any arbitrarily small constant.

Experiments

- **Cricket world cups:** Skill scores of cricket world cup tournaments are decreasing over time.
- **Soccer world cups:** Soccer world cups have remained quite unpredictable over the years.
- **Soccer leagues in 2018-2019:** Recover ranking of soccer leagues that is consistent with fan experience.
- **US mutual funds:** Skill score is minimum during the *Great Recession* of 2008.

Figure: Plots (a), (b), (c), (g), and (h) illustrate *estimated skill PDFs*, and plots (d), (e), (f), and (i) depict corresponding *negative differential entropies*.



(a) Cricket World Cups (b) Soccer World Cups (c) Soccer leagues

(d) Cricket World Cups (e) Soccer World Cups (f) Soccer leagues

(g) US mutual funds (h) US mutual funds (i) US mutual funds

References

[1] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs. I. The method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, December 1952.

[2] S. Negahban, S. Oh, and D. Shah, "Rank centrality: Ranking from pairwise comparisons," *Operations Research, INFORMS*, vol. 65, no. 1, pp. 266–287, January-February 2017.