The Player Kernel

Lucas Maystre, Victor Kristof, Antonio González Ferrer, Matthias Grossglauser
School of Computer and Communication Sciences, EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

MLSA workshop @ ECML-PKDD — September 19th, 2016

Context

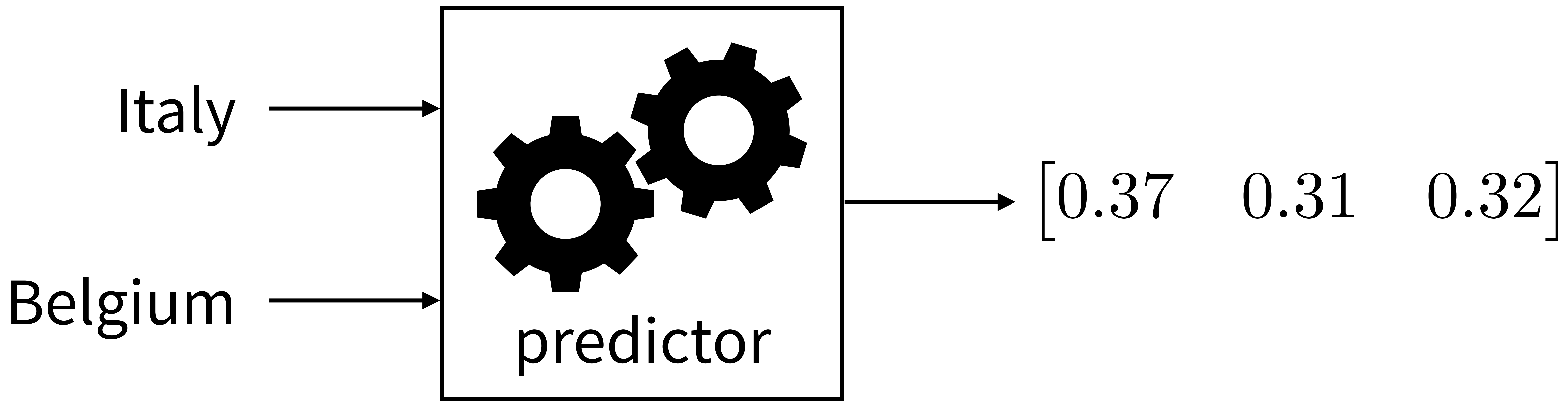
Our entry to the **EURO 2016 Prediction Competition**, Challenge 1

Task: **probabilistic prediction of match outcomes**

Context

Our entry to the **EURO 2016 Prediction Competition**, Challenge 1

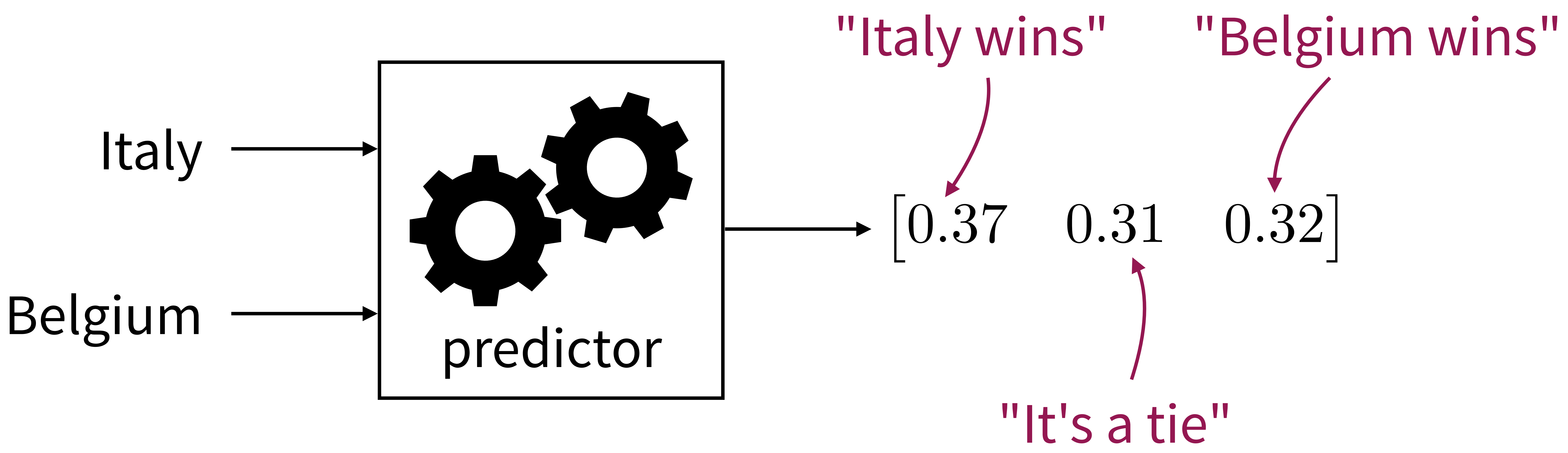
Task: **probabilistic prediction of match outcomes**



Context

Our entry to the **EURO 2016 Prediction Competition**, Challenge 1

Task: **probabilistic prediction of match outcomes**



Starting point

M national teams. Team u has "strength" s_u

Starting point

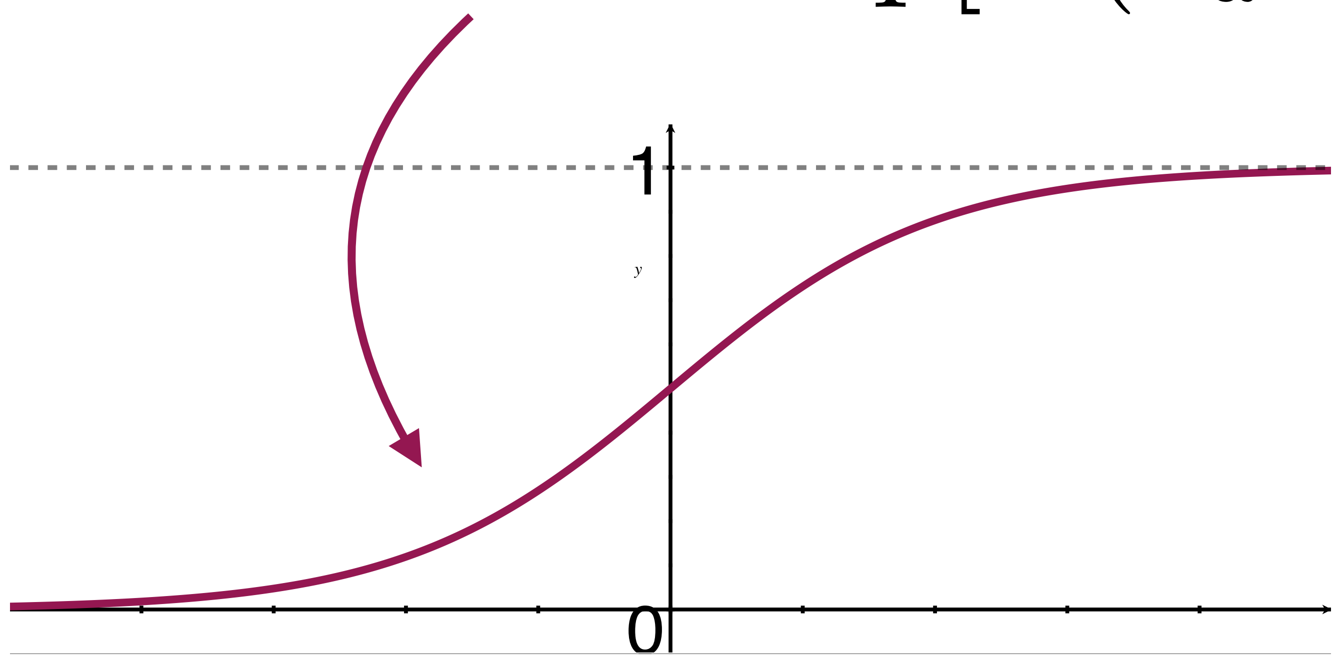
M national teams. Team u has "strength" s_u

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]}$$

Starting point

M national teams. Team u has "strength" s_u

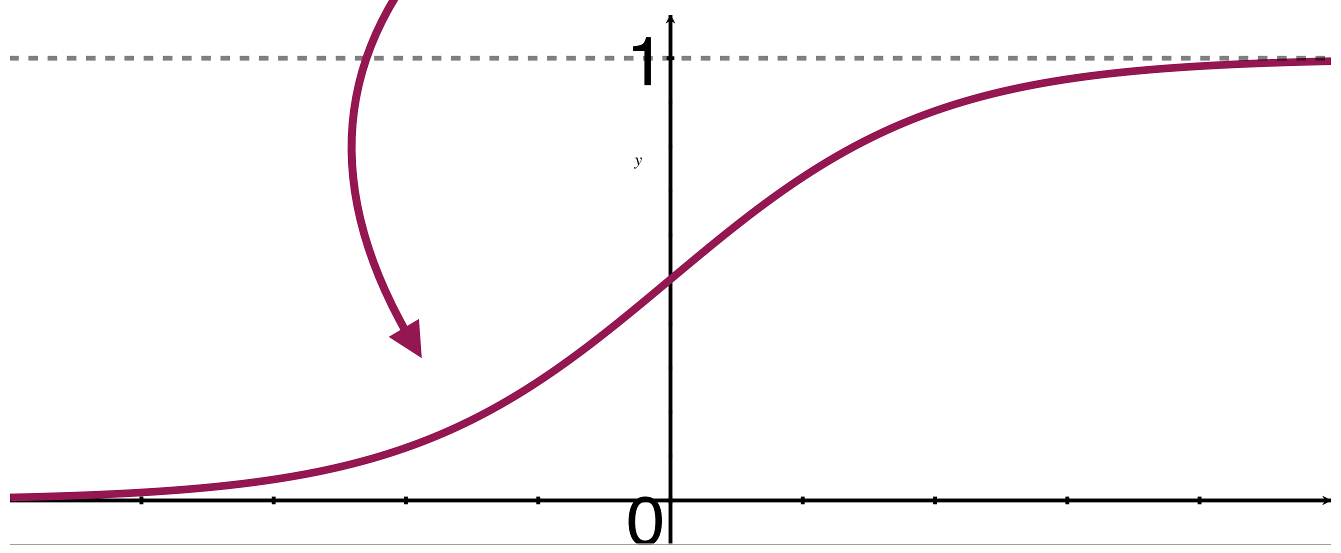
$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]}$$



Starting point

M national teams. Team u has "strength" s_u

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]} = \frac{1}{1 + \exp(-\boldsymbol{s}^\top \boldsymbol{x})}$$

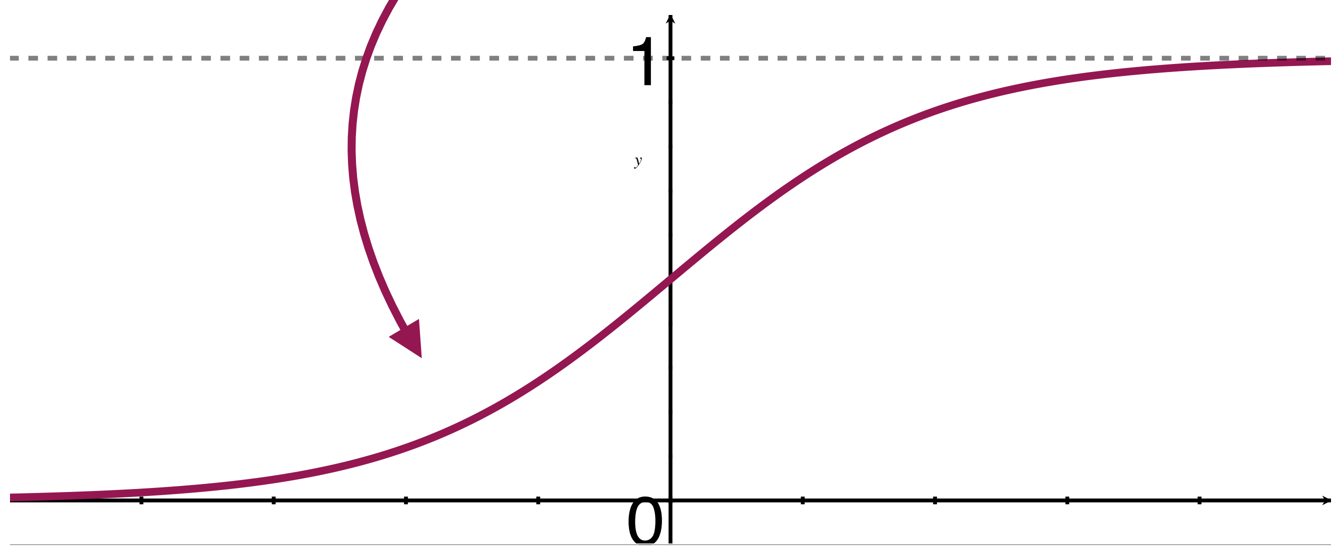


$$\begin{bmatrix} s_1 \\ \vdots \\ s_M \end{bmatrix} \qquad \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ -1 \\ \vdots \\ 0 \end{bmatrix}$$

Starting point

M national teams. Team u has "strength" s_u

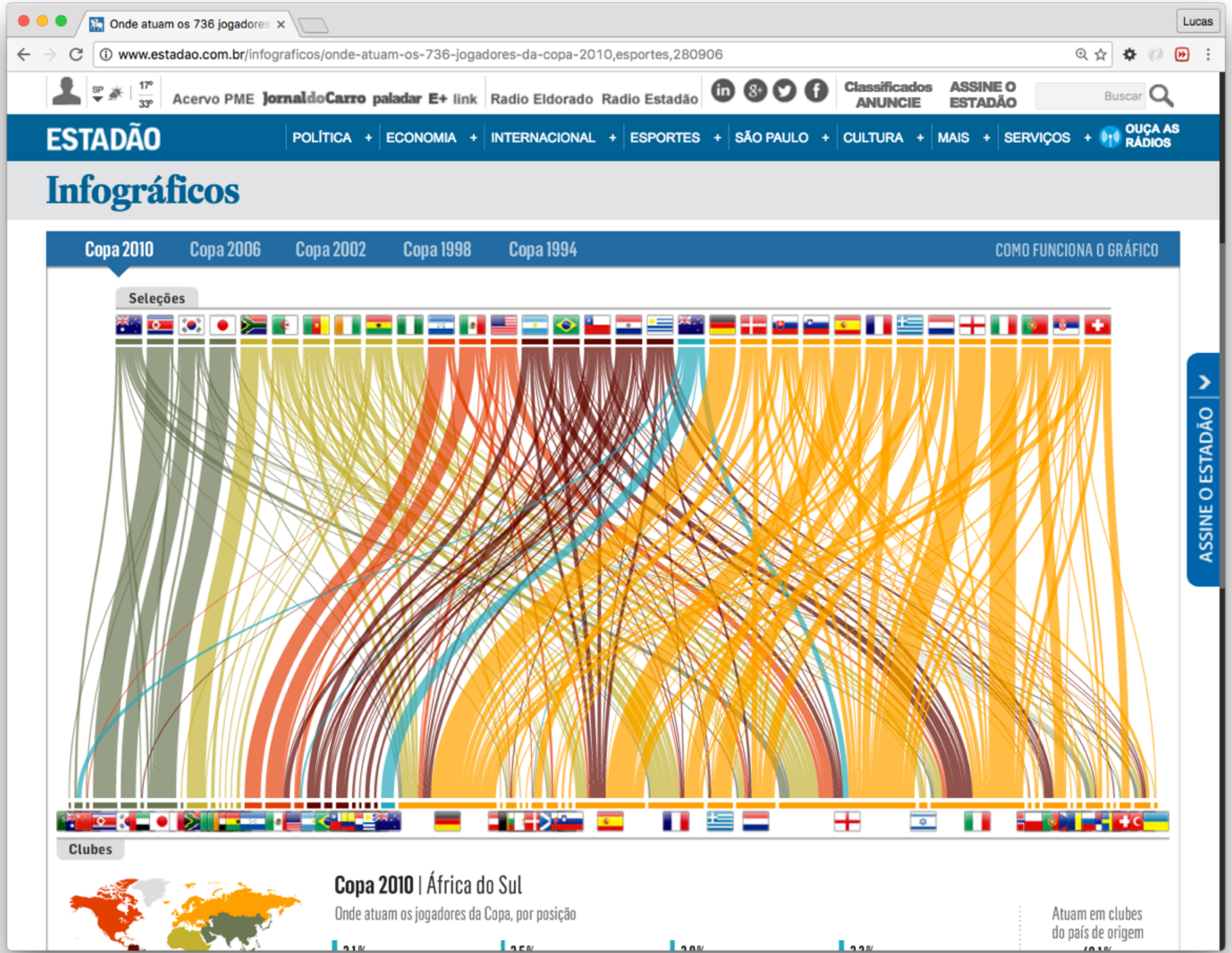
$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]} = \frac{1}{1 + \exp(-\boldsymbol{s}^\top \boldsymbol{x})}$$



$$\boldsymbol{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_M \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ -1 \\ \vdots \\ 0 \end{bmatrix}$$

Key challenges with national teams:

1. They play few matches every year: **recent data is sparse**
2. Their squad change frequently: **old data is stale**



Inspiration

Many players play against each other in **club competitions**

Can we **transfer information** from club matches to international matches?

Source: http://www.estadao.com.br/infograficos/onde-atuam-os-736-jogadores-da-copa-2010,esportes,280906

Main idea

Embed teams in the **space of players.** $s_u = \sum_{i \in \mathcal{L}_u} \tilde{s}_i$

Main idea

Embed teams in the **space of players.** $s_u = \sum_{i \in \mathcal{L}_u} \tilde{s}_i$

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]}$$

Main idea

Embed teams in the **space of players.** $s_u = \sum_{i \in \mathcal{L}_u} \tilde{s}_i$

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]} = \frac{1}{1 + \exp(-\tilde{\boldsymbol{s}}^\top \boldsymbol{z})}$$

$$\tilde{\boldsymbol{s}} = \begin{bmatrix} \tilde{s}_1 \\ \vdots \\ \vdots \\ \tilde{s}_P \end{bmatrix} \qquad \boldsymbol{z} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ -1 \\ -1 \\ -1 \\ \vdots \\ 0 \end{bmatrix}$$

Main idea

Embed teams in the **space of players.** $s_u = \sum_{i \in \mathcal{L}_u} \tilde{s}_i$

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]} = \frac{1}{1 + \exp(-\tilde{\boldsymbol{s}}^\top \boldsymbol{z})}$$

$$\tilde{\boldsymbol{s}} = \begin{bmatrix} \tilde{s}_1 \\ \vdots \\ \vdots \\ \tilde{s}_P \end{bmatrix} \qquad \boldsymbol{z} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ -1 \\ -1 \\ -1 \\ \vdots \\ 0 \end{bmatrix}$$

Club matches and international matches **share the same parameters**.

Main idea

Embed teams in the **space of players.** $s_u = \sum_{i \in \mathcal{L}_u} \tilde{s}_i$

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v)]} = \frac{1}{1 + \exp(-\tilde{\boldsymbol{s}}^\top \boldsymbol{z})}$$

$$\tilde{\boldsymbol{s}} = \begin{bmatrix} \tilde{s}_1 \\ \vdots \\ \vdots \\ \tilde{s}_P \end{bmatrix} \qquad \boldsymbol{z} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ -1 \\ -1 \\ -1 \\ \vdots \\ 0 \end{bmatrix}$$

Club matches and international matches **share the same parameters.**

The number of parameters **explodes.**

Seems like it will lead to **statistical** and **computational** issues.

Bayesian approach

Keep a **distribution over parameters** instead of optimizing the likelihood.

$$\underbrace{p(\tilde{\boldsymbol{s}} \mid \mathcal{D})}_{\text{posterior distribution}} \propto \underbrace{p(\mathcal{D} \mid \tilde{\boldsymbol{s}})}_{\text{likelihood}} \times \underbrace{p(\tilde{\boldsymbol{s}})}_{\substack{\text{prior distribution} \\ \text{(e.g. Gaussian)}}}$$

Bayesian approach

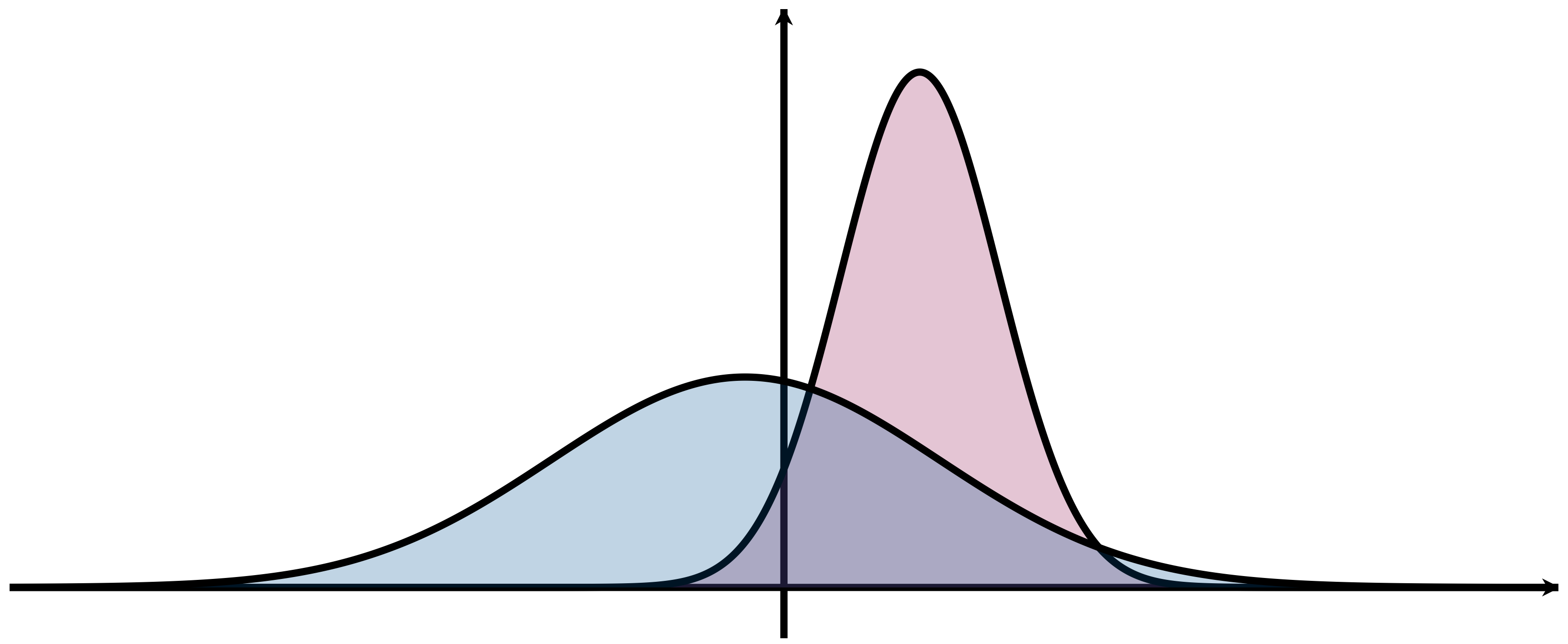
Keep a **distribution over parameters** instead of optimizing the likelihood.

$$p(\tilde{\boldsymbol{s}} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \tilde{\boldsymbol{s}}) \times p(\tilde{\boldsymbol{s}})$$

posterior distribution

likelihood

prior distribution
(e.g. Gaussian)

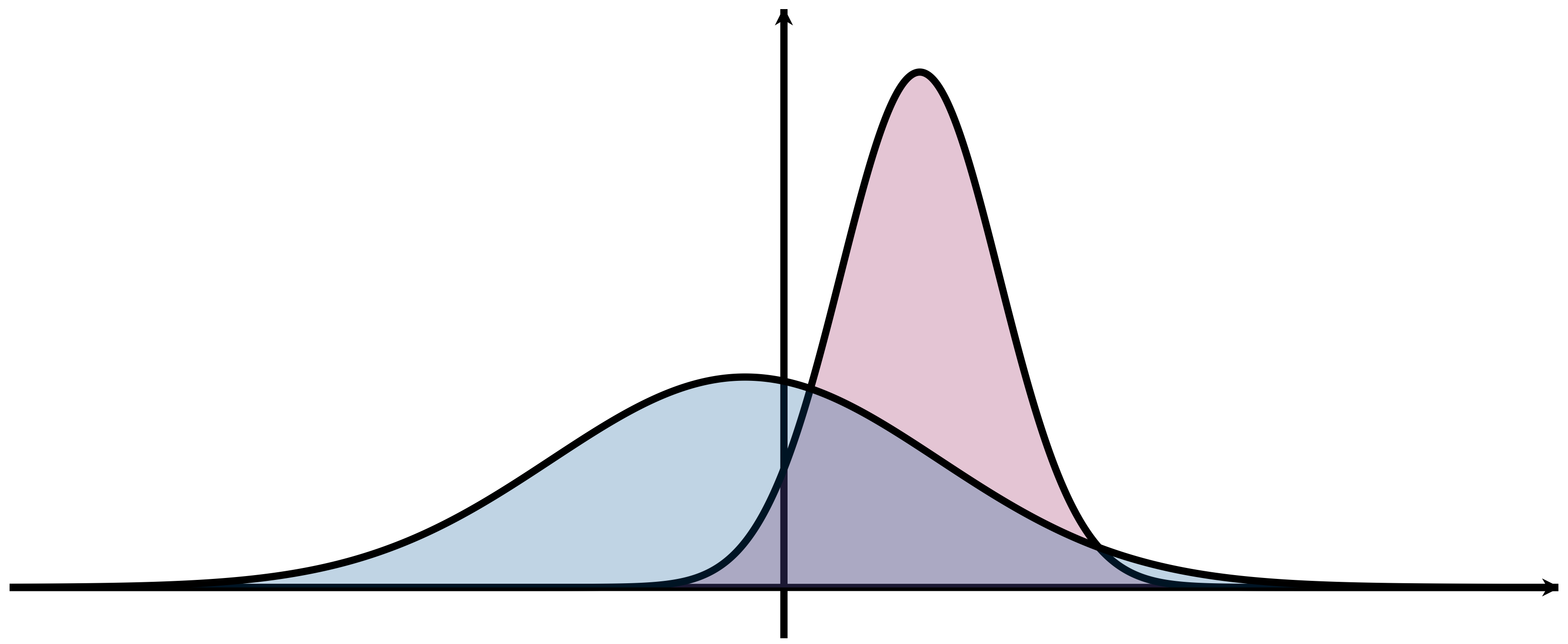


Bayesian approach

Keep a **distribution over parameters** instead of optimizing the likelihood.

$$\underbrace{p(\tilde{\boldsymbol{s}} \mid \mathcal{D})}_{\text{posterior distribution}} \propto \underbrace{p(\mathcal{D} \mid \tilde{\boldsymbol{s}})}_{\text{likelihood}} \times \underbrace{p(\tilde{\boldsymbol{s}})}_{\text{prior distribution}}$$

(e.g. Gaussian)



Statistical issues solved? **not clear**
Computational issues solved? **no**

Dual viewpoint

In fine, we are **only interested in** $p(\tilde{\boldsymbol{s}}^\top \boldsymbol{z} \mid \mathcal{D})$ → "strength" difference

Dual viewpoint

In fine, we are **only interested in** $p(\tilde{\boldsymbol{s}}^\top \boldsymbol{z} \mid \mathcal{D})$ → "strength" difference

Accurate estimation of all parameters is not necessary

Dual viewpoint

In fine, we are **only interested in** $p(\tilde{\boldsymbol{s}}^\top \boldsymbol{z} \mid \mathcal{D})$ → "strength" difference

Accurate estimation of all parameters is not necessary

$$f(\boldsymbol{z}) = \tilde{\boldsymbol{s}}^\top \boldsymbol{z} \longrightarrow p(\boldsymbol{f} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{f}) \times p(\boldsymbol{f})$$

Dual viewpoint

In fine, we are **only interested in** $p(\underline{\tilde{\boldsymbol{s}}^\top \boldsymbol{z}} \mid \mathcal{D})$ → "strength" difference

Accurate estimation of all parameters is not necessary

$$f(\boldsymbol{z}) = \tilde{\boldsymbol{s}}^\top \boldsymbol{z} \longrightarrow p(\boldsymbol{f} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{f}) \times \underline{p(\boldsymbol{f})}$$

$$\mathrm{Cov}[f(\boldsymbol{z}), f(\boldsymbol{z}')] = \sigma^2 \boldsymbol{z}^\top \boldsymbol{z}'$$

Dual viewpoint

In fine, we are **only interested in** $p(\tilde{\boldsymbol{s}}^\top \boldsymbol{z} \mid \mathcal{D})$ → "strength" difference

Accurate estimation of all parameters is not necessary

$$f(\boldsymbol{z}) = \tilde{\boldsymbol{s}}^\top \boldsymbol{z} \longrightarrow p(\boldsymbol{f} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{f}) \times p(\boldsymbol{f})$$

$$\mathrm{Cov}[f(\boldsymbol{z}), f(\boldsymbol{z}')] = \sigma^2 \boldsymbol{z}^\top \boldsymbol{z}'$$

Inference can be done in the dual space

Dual viewpoint

In fine, we are **only interested in** $p(\tilde{\boldsymbol{s}}^\top \boldsymbol{z} \mid \mathcal{D})$ → "strength" difference

Accurate estimation of all parameters is not necessary

$$f(\boldsymbol{z}) = \tilde{\boldsymbol{s}}^\top \boldsymbol{z} \longrightarrow p(\boldsymbol{f} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{f}) \times p(\boldsymbol{f})$$

$$\mathrm{Cov}[f(\boldsymbol{z}), f(\boldsymbol{z}')] = \sigma^2 \boldsymbol{z}^\top \boldsymbol{z}'$$

Inference can be done in the dual space

$$f(\boldsymbol{z}) \sim \mathcal{GP}[\boldsymbol{0}, k(\boldsymbol{z}, \boldsymbol{z}')]$$

Dual viewpoint

In fine, we are **only interested in** $p(\tilde{\boldsymbol{s}}^\top \boldsymbol{z} \mid \mathcal{D})$ → "strength" difference

Accurate estimation of all parameters is not necessary

$$f(\boldsymbol{z}) = \tilde{\boldsymbol{s}}^\top \boldsymbol{z} \longrightarrow p(\boldsymbol{f} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{f}) \times p(\boldsymbol{f})$$

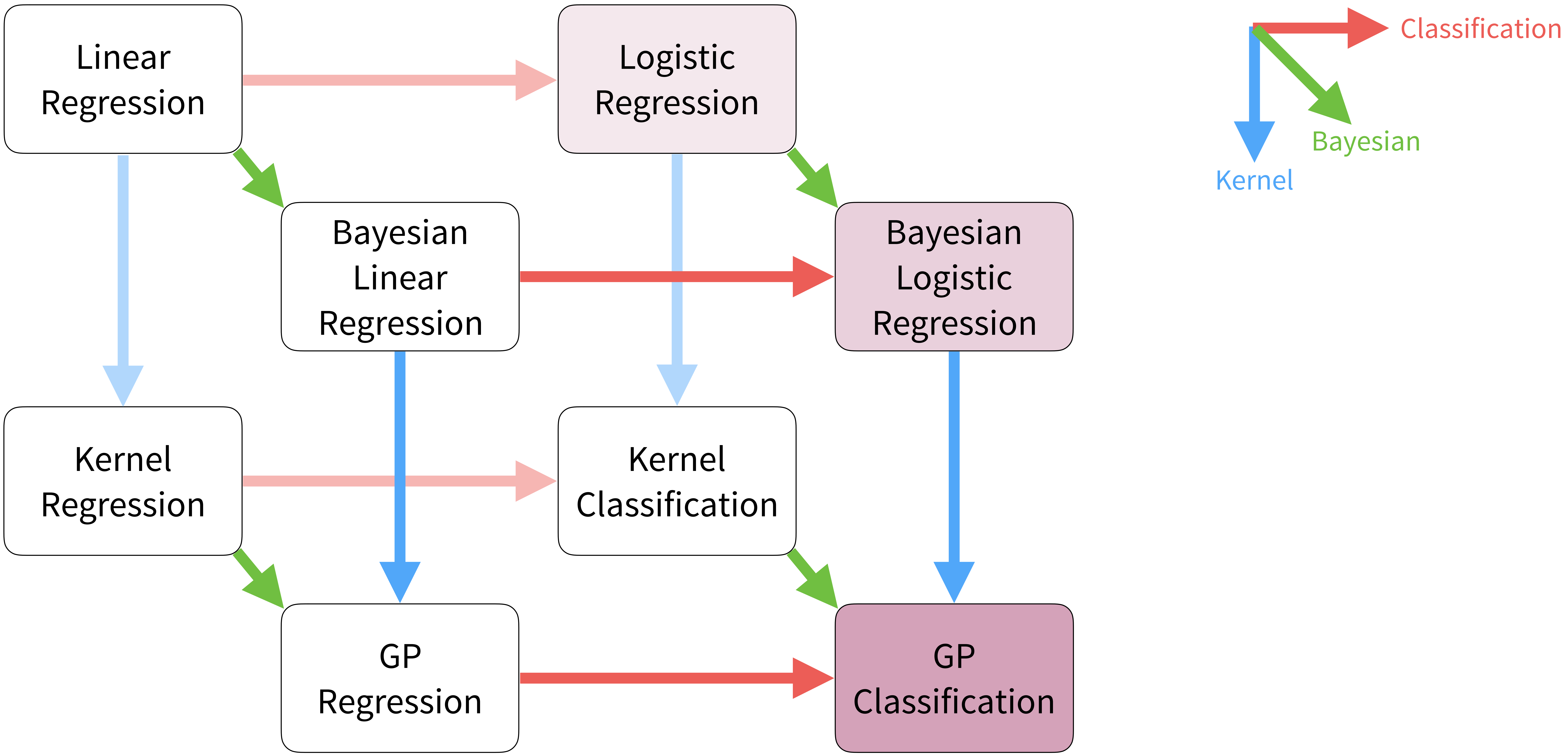
$$\mathrm{Cov}[f(\boldsymbol{z}), f(\boldsymbol{z}')] = \sigma^2 \boldsymbol{z}^\top \boldsymbol{z}'$$

Inference can be done in the dual space

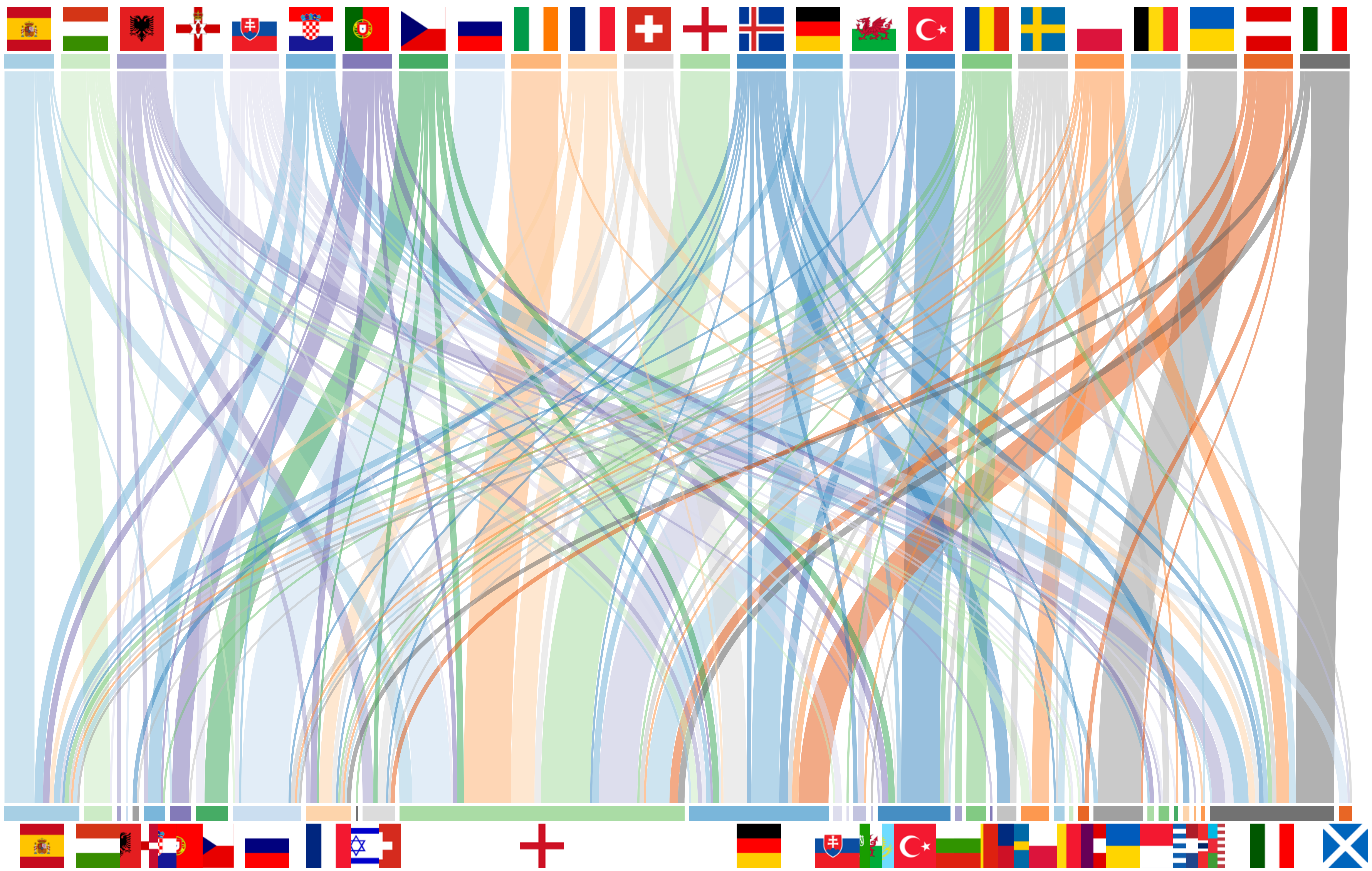
$$f(\boldsymbol{z}) \sim \mathcal{GP}[\mathbf{0}, k(\boldsymbol{z}, \boldsymbol{z}')]$$

→ **The player kernel!**

The cube



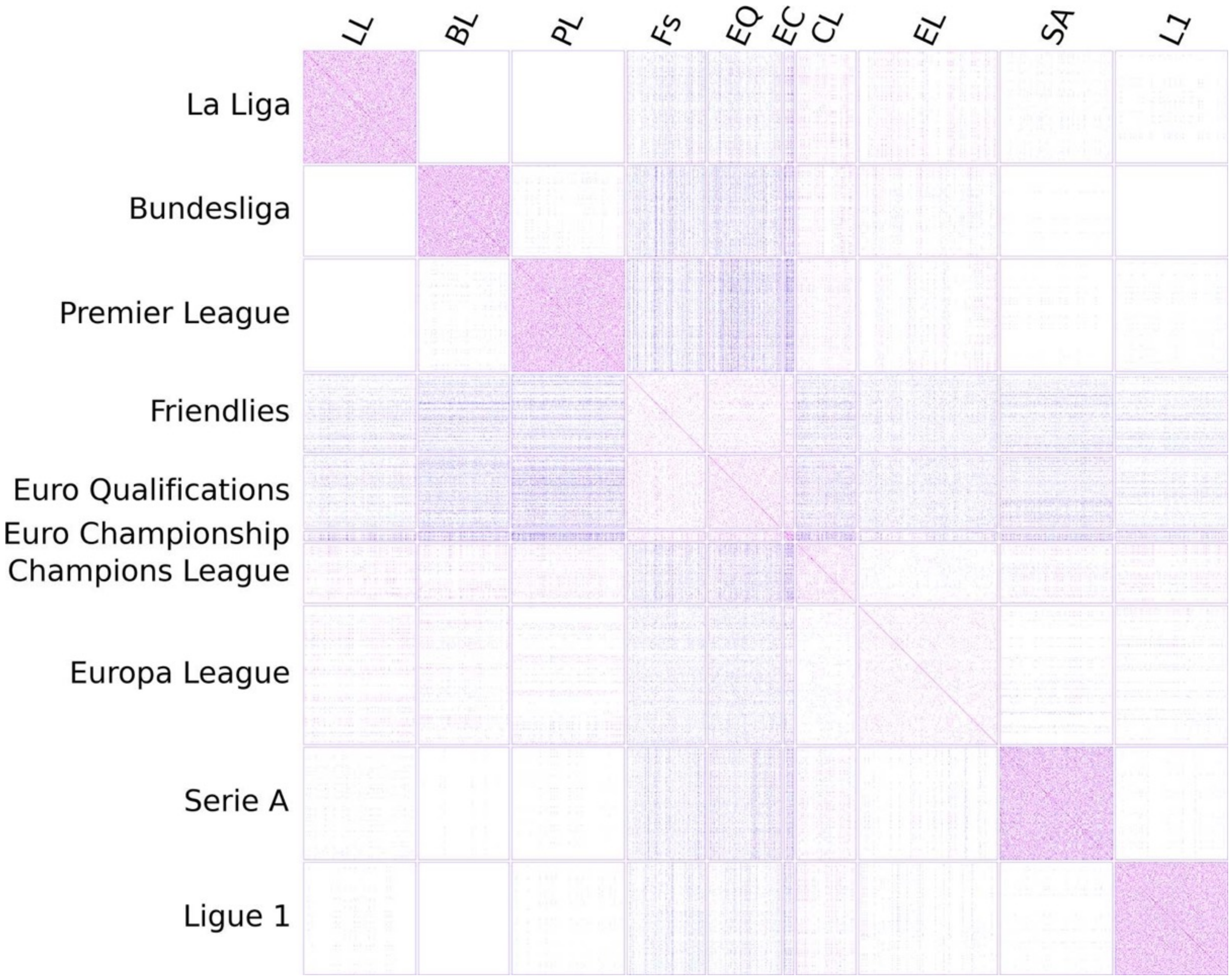
Credit: Zoubin Ghahramani



Dataset

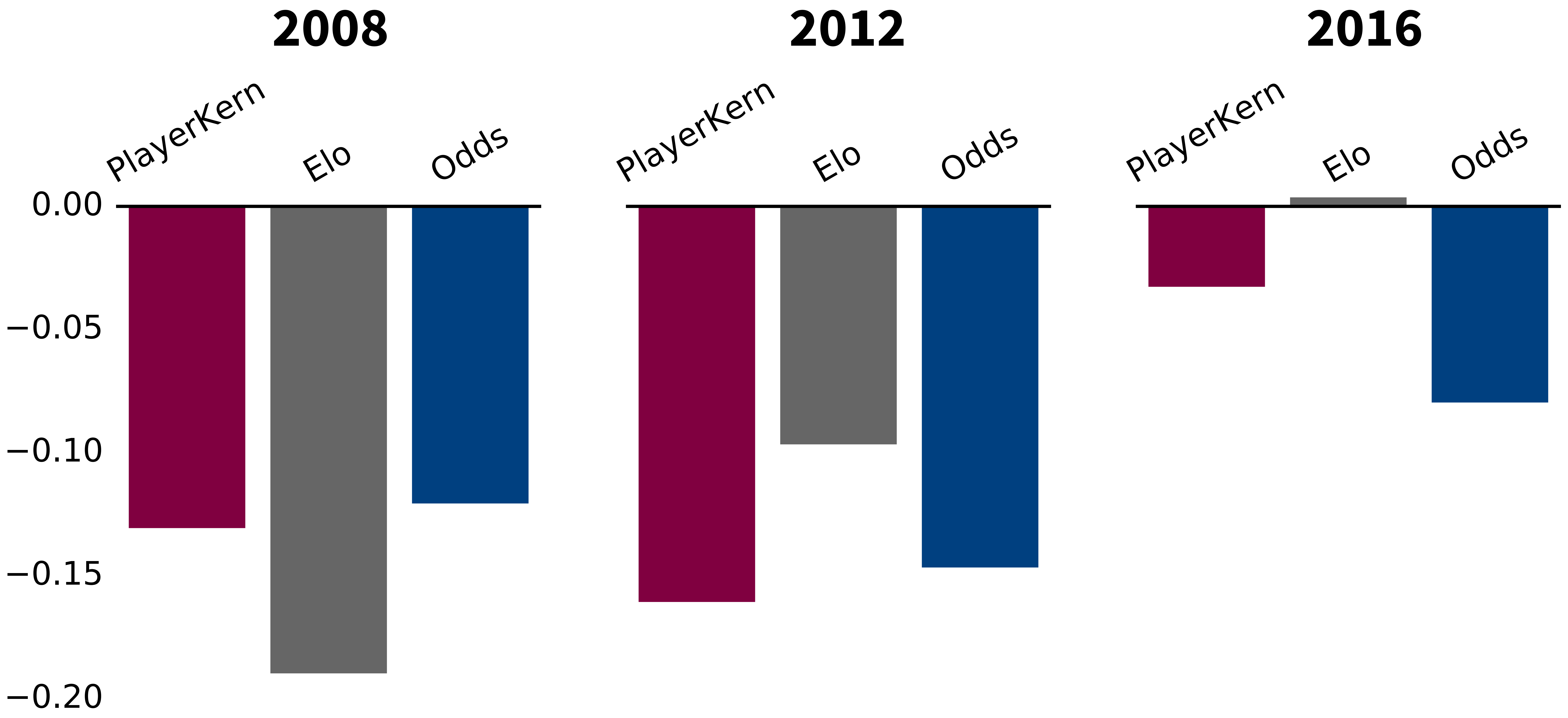
Collected data on **24 887 matches** from main football competitions over the last **10 years**.

33 157 distinct players appear in the dataset.

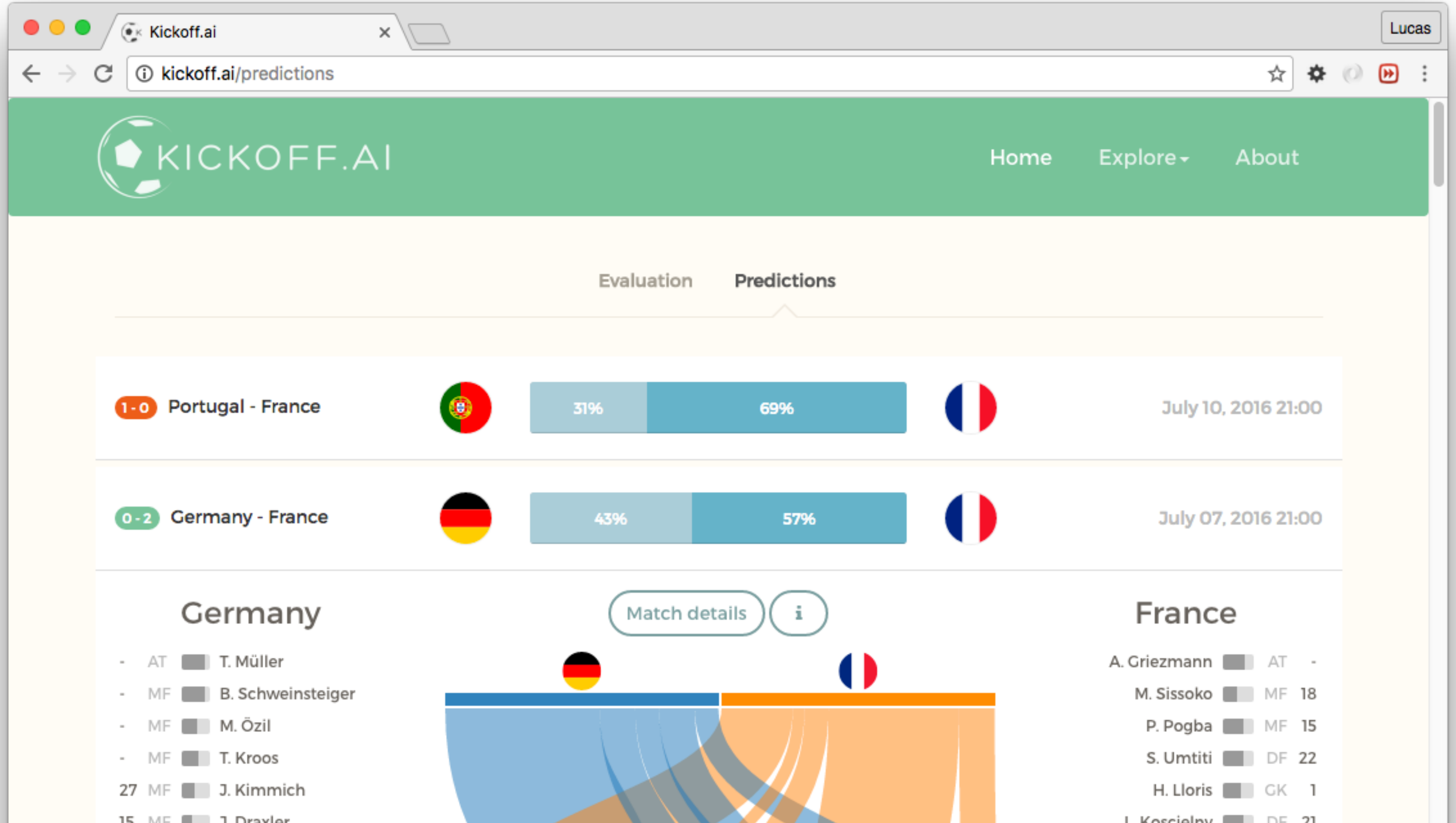

LL
BL
PL
Fs
EQ
EC
CL
EL
SA
L1
La Liga
Bundesliga
Premier League
Friendlies
Euro Qualifications
Euro Championship
Champions League
Europa League
Serie A
Ligue 1

Results

Logarithmic loss against competing approaches in 2008, 2012 and 2016.



KICKOFF.AI



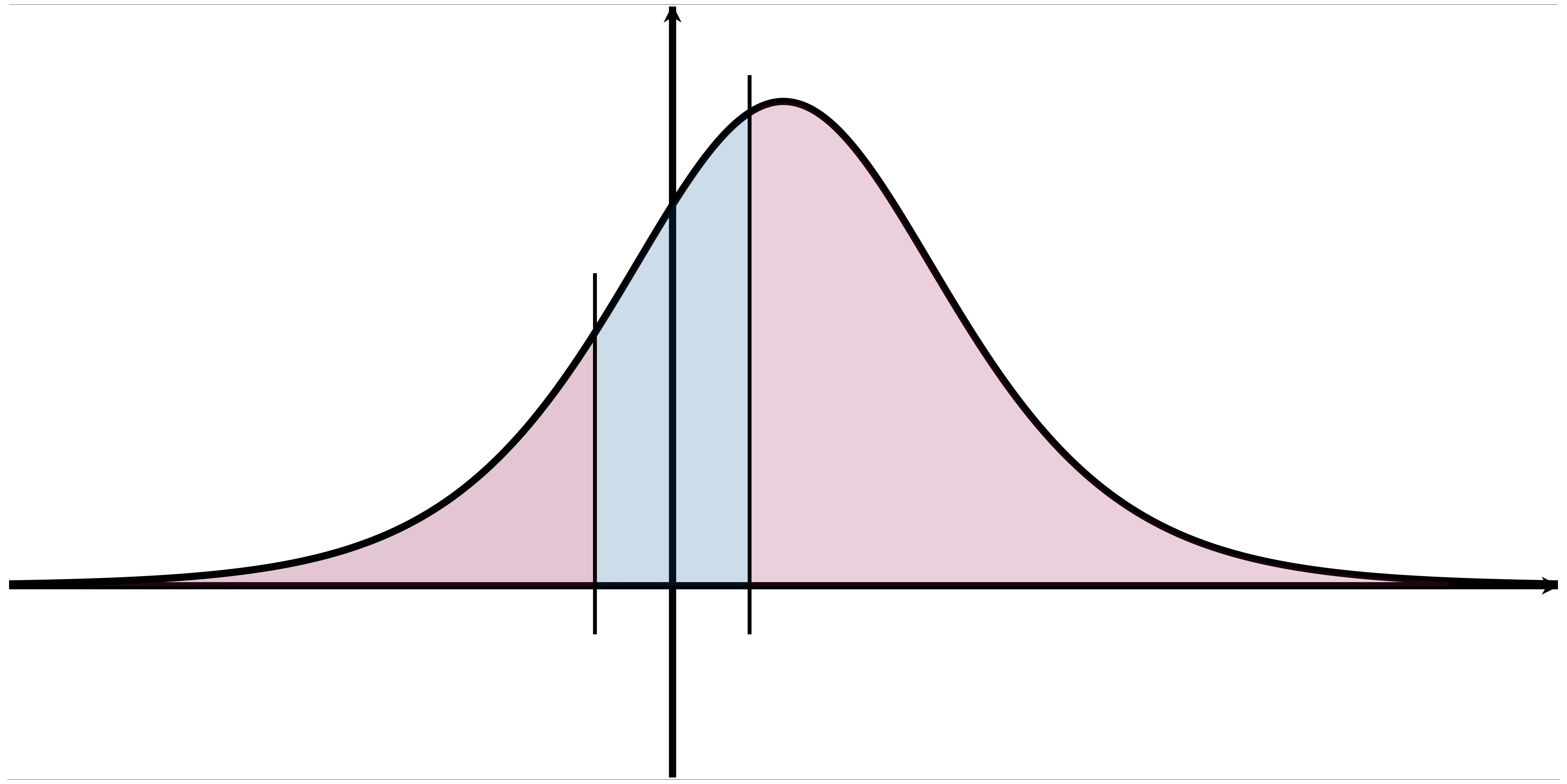
Kickoff.ai
Lucas
kickoff.ai/predictions
KICKOFF.AI
Home
Explore
About
Evaluation
Predictions
1 - 0
Portugal - France
31%
69%
July 10, 2016 21:00
0 - 2
Germany - France
43%
57%
July 07, 2016 21:00
Germany
Match details
France
- AT T. Müller
- MF B. Schweinsteiger
- MF M. Özil
- MF T. Kroos
27 MF J. Kimmich
A. Griezmann AT -
M. Sissoko MF 18
P. Pogba MF 15
S. Umtiti DF 22
H. Lloris GK 1

Ternary outcomes

Rao and Kupper (1967) proposed the following extension.

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v - \alpha)]}$$

$$\begin{aligned} P(u \equiv v) &= 1 - P(u \succ v) - P(v \succ u) \\ &\propto P(u \succ v) \cdot P(v \succ u) \end{aligned}$$

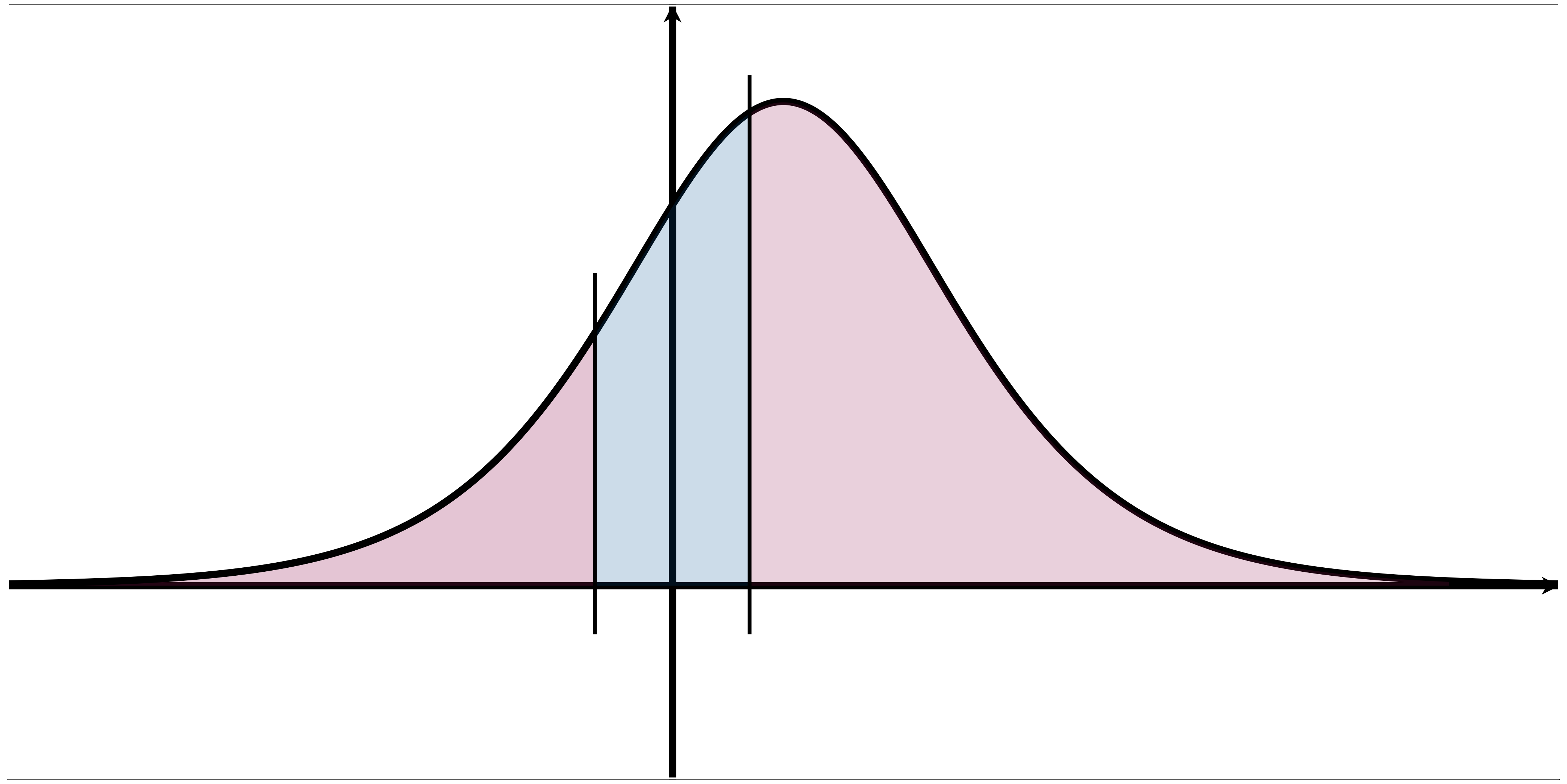


Ternary outcomes

Rao and Kupper (1967) proposed the following extension.

$$P(u \succ v) = \frac{1}{1 + \exp[-(s_u - s_v - \alpha)]}$$

$$P(u \equiv v) = 1 - P(u \succ v) - P(v \succ u)$$
$$\propto P(u \succ v) \cdot P(v \succ u)$$



A draw is (essentially) equivalent to **one win** and **one loss**.