



HAL
open science

Sur la construction et le pouvoir prédictif du classement ELO

Paul Steffen, Léo Gerville-Réache

► **To cite this version:**

Paul Steffen, Léo Gerville-Réache. Sur la construction et le pouvoir prédictif du classement ELO. 52ème Journées de statistique de la société Française de statistique (SFDS), 2020, Nice, France. p 725-730. hal-02983379

HAL Id: hal-02983379

<https://hal.science/hal-02983379v1>

Submitted on 17 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SUR LA CONSTRUCTION ET LE POUVOIR PRÉDICTIF DU CLASSEMENT ELO

Paul Steffen ^{1,2} & Léo Gerville-Réache ²

¹ *Betclik 117 Quai de Bacalan, 33300 Bordeaux, p.steffen@betcligroup.com*

² *Univ. Bordeaux, CNRS, IMS, UMR 5218, 33405 Talence,
leo.gerville-reache@u-bordeaux.fr*

Résumé. Dans de nombreuses disciplines sportives, les classements sont utilisés comme un outil de comparaison entre les participants à une compétition. Le classement Elo, initialement utilisé dans les échecs, a prouvé par ses nombreuses applications qu'il parvenait à donner une image fidèle du niveau de jeu des compétiteurs. Dans cette communication, nous utilisons les adaptations de ce système de classement, pour prédire l'issue des rencontres de football, de tennis et de basketball.

Mots-clés. Elo, classement, notation, prédiction statistique des résultats sportifs

Abstract. In many sports, rankings are used as a comparison tool between participants in a competition. The Elo rating, originally used in chess, has proven by its many applications that it gives a true picture of the competitors' strength. In this communication, we use adaptations of this rating system to predict the outcome of football, tennis and basketball matches.

Keywords. Elo, ranking, rating, sports forecasting

1 Introduction

En 1960, la Fédération américaine des échecs adopte la procédure de classement des joueurs proposée par Arpad Elo. Cette méthode de classement rend possible une évaluation quantitative du niveau de jeu des participants à une compétition. Utilisé depuis pour le jeu de go ou encore le e-sport, le classement Elo a été aussi employé pour classer les photos de profil des membres de la 1ère version de The Facebook.

En sport, chaque fédération utilise une méthode de classement spécifique. Les méthodes évoluent régulièrement et sont souvent éloignées de la proposition d'Arpad Elo.

Dans cette communication, nous revenons sur les principes du classement Elo et questionnons son emploi dans le cadre de l'estimation des probabilités de victoire, de défaite et de nul d'une rencontre sportive. Aussi, après avoir rappelé les mécanismes du classement Elo et ses adaptations aux spécificités sportives, nous appliquons cette procédure au tennis, basket et football et discutons de la pertinence d'un tel classement pour estimer les probabilités des différentes issues possibles d'une rencontre sportive.

2 Le classement, un indicateur de niveau de jeu

Dans le monde du sport, les classements peuvent avoir de nombreux objectifs. De manière évidente, ces derniers permettent d'obtenir une indication sur le niveau actuel et la progression des compétiteurs. Ils permettent également aux fédérations de motiver leurs licenciés à participer à des compétitions. Le tennis, par exemple, utilise son propre système de classement, dans les 2 associations professionnelles que sont l'ATP¹ et le WTA².

2.1 Le classement Elo

Basé sur les performances antérieures des compétiteurs, ce système de classement assigne à chaque participant un nombre de points, tel que 2 compétiteurs ayant le même niveau de jeu ont le même nombre de points. Plus un compétiteur a un niveau élevé, plus il aura de points. Mis à part des ajustements, spécifiques aux cas d'application, ce classement est un jeu à somme nulle où plus les adversaires d'une rencontre ont une différence de points importante, plus le nombre de points échangés en cas de victoire du favori sera faible et inversement si l'outsider remporte la partie. Ainsi, le score Elo moyen d'un ensemble de joueurs en compétition est égal à la valeur initial du système de classement.

Dans ce classement, la performance d'un compétiteur est modélisée par une variable aléatoire centrée sur la notation de ce dernier. Initialement, la distribution de cette variable suivait une loi normale, avant de suivre une loi logistique permettant des queues de distribution plus lourdes.

Les points des participants à une rencontre sont mis à jour à l'issue de leur opposition de la manière suivante:

$$elo'_i = elo_i + K \cdot G(O - E) \quad (1)$$

où l'issue espérée de la rencontre est E et le résultat observé est $O \in \{0, 0.5, 1\}$, correspondant respectivement à la défaite, au match nul et à la victoire.

Le paramètre K est un paramètre d'échelle concernant l'impact des résultats les plus récents sur le Elo. Ce paramètre déterminant la volatilité du classement, peut être augmenté pour les joueurs ayant disputé peu de rencontres, afin d'entraîner des changements plus importants et une progression vers le niveau réel plus rapide.

Le paramètre G est aussi un paramètre d'échelle. Ce dernier n'apparaît pas dans la forme initiale de la formule d'actualisation des points Elo. Il a été ajouté afin de s'adapter aux spécificités des différents cas d'application de ce système de classement. Il peut permettre d'identifier l'importance d'une rencontre, ou une marge de victoire importante afin d'augmenter le nombre de points échangés lors de la rencontre.

Le paramètre E , d'issue espérée, s'exprime à l'aide de la relation suivante:

$$E = \frac{1}{1 + 10^{\frac{-dr_A}{400}}} \quad (2)$$

¹Association of Tennis Professionals

²Women's Tennis Association

où $dr_A = elo_A - elo_B$ est la différence de points entre les 2 compétiteurs. Plus ce résultat sera proche de 1, plus l'équipe A sera favorite, et inversement.

2.2 Les spécificités associées aux sports étudiés

Créé par Arpad Elo, ce classement fût imaginé pour les échecs et son efficacité a encouragé son utilisation dans de tout autre domaines.

Ainsi, les applications concernant des sports comme le football ou le basket ont intégré l'avantage induit par le fait de jouer à domicile, ou l'augmentation de points échangés pour une victoire avec une marge importante. Concernant le tennis, des ajustements ont été effectués pour l'importance du match, et le type de surface sur lequel le match est joué.

On différenciera alors les spécificités influençant l'issue de la rencontre (2), de celles n'intervenant que lors de l'évolution des points Elo suite à une rencontre (1).

L'avantage à domicile, présent dans le football et le basketball, sera incorporé par l'ajout d'une constante dans la différence des points des opposants, en faveur de l'équipe à domicile, de la manière suivante: $dr_A = elo_A - elo_B + HomeAdvantage$.

Pour le tennis, l'ajustement concernant la surface sur laquelle se dispute le match, se manifeste par une moyenne pondérée entre les points Elo spécifiques à la surface, et les points Elo de toutes les surfaces: $elo = \lambda \cdot elo_{global} + (1 - \lambda) \cdot elo_{surface}$.

Enfin, l'amplification de l'évolution des points selon la différence de score au football s'effectuera de la manière suivante: si $\Delta_{goals} \leq 1$, alors $G = 1$, si $\Delta_{goals} = 2$, alors $G = 1,5$, sinon $G = \frac{11 + \Delta_{goals}}{8}$. Pour le basketball, Silver, N. et Fischer-Baum R. (2015) proposent $G = \frac{(\Delta_{score} + 3)^{0.8}}{7.5 + 0.006 \cdot dr_A}$.

Concernant le tennis, G sera influencé par l'importance du match à l'aide d'un coefficient, et selon le classement du joueur à l'aide de la formule suivante proposée par Cekovic, M. (2015): $G = 1 + \frac{18}{1 + 2^{(rating - 1500)/63}}$

3 Prédiction de l'issue d'une rencontre

Un système de classement comme celui du Elo représente alors un double intérêt pour une modélisation statistique de l'issue d'une rencontre: il peut être utilisé directement pour modéliser la probabilité d'issue de la rencontre, ou pour générer une variable, représentant le niveau de performance du compétiteur, utile à une modélisation plus complète, utilisant d'autres variables.

3.1 Du classement Elo à la probabilité de résultat

Bien que la formule d'actualisation des points Elo (1) intègre la possibilité d'un match nul ($\theta = 0,5$), l'éventualité que cet événement survienne n'apparaît pas dans la relation

(2).

Afin de construire la probabilité de match nul, possible lors d'un match de football, la densité de probabilités empirique de l'issue des rencontres des 5 championnats majeurs européens de 1993 à 2020 a été utilisée. Ces matchs ont été ordonnés selon la différence de cotes du bookmaker Bet365 entre la victoire de l'équipe à domicile et celle de l'équipe à l'extérieur dans l'histogramme groupé de la Figure 1.

La modélisation de ces résultats empiriques, visibles sur les courbes de la Figure 1, a été effectuée à l'aide d'un classement par régression logistique ordinaire inspiré de celui proposé par Lasek, J. (2016) où les probabilités d'occurrence des événements sont déterminées de la manière suivante:

$$\begin{aligned}
 P(A) &= \frac{1}{1 + 10^{c - \frac{drA}{400}}} & P(B) &= 1 - \frac{1}{1 + 10^{-c - \frac{drA}{400}}} \\
 P(draw) &= \frac{1}{1 + 10^{-c - \frac{drA}{400}}} - \frac{1}{1 + 10^{c - \frac{drA}{400}}} & & (3)
 \end{aligned}$$

où $c \in R_+$ est le paramètre déterminant la part des matchs nuls, évalué par une minimisation de l'entropie croisée, tout comme le *Home Advantage* incorporé à la différence de points Elo.

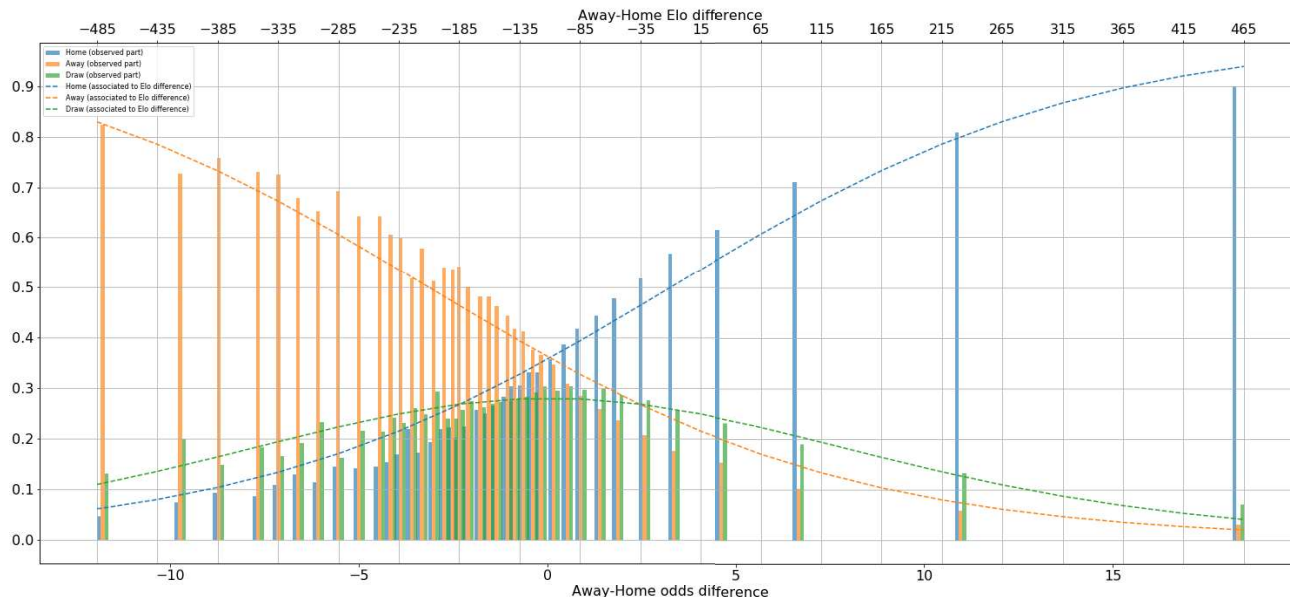


Figure 1: Part des issues des rencontres de football des 5 championnats majeurs européens de 1993 à 2020 et probabilités d'issue associées à une différence de points Elo

4 Application aux principaux championnats de chaque discipline

A l'aide des résultats des rencontres des 5 championnats majeurs européens de football depuis la saison 2009/2010, des matchs du championnat de basketball NBA depuis la saison 2000/2001, et des résultats du circuit de tennis ATP depuis 1991, 3 systèmes de classement Elo ont été évalués.

4.1 Méthode d'évaluation

Dans un objectif d'estimer les probabilités des différentes issues possibles d'une rencontre sportive, 2 métriques d'évaluation sont utilisées. La principale est la perte logistique, permettant de prendre en compte l'incertitude d'une prédiction à l'aide de sa différence avec le label observé:

$$L_{\log}(Y, P) = -\log \Pr(Y|P) = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} y_{i,k} \log p_{i,k} \quad (4)$$

où N est le nombre de rencontres, K appartient à l'ensemble des labels d'issue de la rencontre $\{Home, Away, Draw\}$, $p_{i,k}$ est la probabilité d'issue de la rencontre, calculée à l'aide des relations (3.1) et $y_{i,k} = 1$ si le résultat observé a le label k et 0 sinon.

Enfin, l'exactitude, définie comme le nombre de prédictions dont la valeur prévue est égale à la valeur réelle, permet une meilleure interprétabilité des résultats.

Ces métriques ont été calculées sur l'ensemble des rencontres de la saison 2017/2018, pour les 3 sports.

4.2 Résultats

Afin de mettre en perspective les résultats d'une prédiction à l'aide de la différence de points Elo de 2 opposants lors d'une rencontre sportive, il a été choisi d'utiliser les prédictions issues d'une classification naïve bayésienne, nécessitant comme le Elo, peu de données relatives à la rencontre (équipe à domicile pour le football et le basketball, et joueur le mieux classé au classement ATP pour le tennis).

Model	Log-loss	Accuracy
Soccer Elo Ranking	0.98	0.53
Soccer Naive Bayesian	1.07	0.45
Tennis Elo Ranking	0.63	0.65
Tennis Naive Bayesian	0.66	0.64
Basket Elo Ranking	0.62	0.65
Basket Naive Bayesian	0.68	0.59

Table 1: Métriques d'évaluation calculées sur la saison 2017/2018

4.3 Discussion

A la vue de nos résultats, on remarque un faible gain de performance pour le tennis, s'expliquant par un bon niveau d'information contenu dans le classement ATP afin de déterminer l'issue d'une rencontre.

Dans notre objectif de proposer une prédiction statistique de l'issue d'une rencontre sportive, le Elo va plus loin qu'une simple comparaison des rangs entre les opposants au sein d'un classement. Par l'utilisation de la différence de points Elo, la prédiction est basée sur une comparaison relative de la force des opposants.

Cependant, de nombreux paramètres déterminant cette force (composition initiale d'une équipe, conditions climatiques, style de jeu etc...) ne sont pas utilisés. Et, bien que le système Elo permette une amélioration allant jusqu'à 0.1 point d'exactitude dans la prédiction de l'issue d'une rencontre sportive, par rapport à un modèle naïf, l'utilisation de points Elo semble plus pertinente au sein d'un modèle d'apprentissage statistique plus élaboré.

Bibliographie

Lasek, J. (2016). EURO 2016 Predictions Using Team Rating Systems. MLSA@PKDD/ECML.

Lasek, J. et Szlávik, Z. et Bhulai, S. (2013). The predictive power of ranking systems in association football. Int. J. of Applied Pattern Recognition. 1. ied Pattern Recognition. 10.1504/IJAPR.2013.052339.

EloRatings.net (2012). The World Football Elo Rating System.
<http://www.eloratings.net/about>

Lacy, S. (2018). Implementing an Elo rating system for European football.
<https://stuartlacy.co.uk/2017/08/31/implementing-an-elo-rating-system-for-european-football/>

Silver, N. et Fischer-Baum R. (2015). How we calculate NBA Elo Ratings.
<https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/>

FIFA.com (2018). Revision of the FIFA / Coca-Cola World Ranking.
<http://www.fifa.com/worldranking/procedureandschedule/menprocedure/index.html>

Cekovic, M. (2015). Tennis Crystal Ball.
<https://github.com/mcekovic/tennis-crystal-ball>

FootballDatabase. FootballDatabase methodology. <https://footballdatabase.com/methodology.php>