

Université du Québec
Institut national de la recherche scientifique
Centre Énergie Matériaux Télécommunications

**ÉTUDE ET AMÉLIORATION DU SYSTÈMES DE NOTATION ELO POUR
LES COMPÉTITIONS SPORTIVES**

Par
Djebbi Aymen

Mémoire présenté pour l'obtention du grade de
Maître es Sciences, M.Sc.
en Télécommunications

Jury d'évaluation

Examineur externe	Ismail Ben Ayed ÉTS
Examineur interne	Jean-Charles Grégoire INRS
Directeur de recherche	Leszek Szczecinski INRS

Remerciements

Je voudrais exprimer mes remerciements et ma sincère gratitude envers mon superviseur, le Professeur Leszek Szczecinski. Ses conseils et ses encouragements m'ont été d'une grande aide tout au long de mes études de maîtrise. Sa porte m'était toujours ouverte pour discuter de toutes mes pensées et interrogations. Et je tiens surtout à le remercier pour sa patience et son soutien.

J'aimerais également remercier mes collègues, Hsan, Amen, Rihem et mes autres collègues, pour les moments merveilleux que nous avons partagés ensemble au cours de ces deux dernières années. Leur compagnie a considérablement enrichi mon expérience au sein du centre EMT de l'INRS.

Je profite également de cette occasion pour remercier mes amis d'enfance, Amine, Hatem et Ghassen pour leurs encouragements et leur soutien continu.

Je tiens également à remercier ma fiancée et l'amour de ma vie pour sa patience, son soutien et ses encouragements. Elle était toujours présente pour m'appuyer dans tout ce que j'entreprends.

Finalement, je tiens à remercier ma famille bien-aimée, ma mère, mes frères et ma sœur. C'est leur soutien et encouragements continus qui ont rendu ce travail possible. Je dédie ce travail spécialement à mon défunt père l'idole de ma vie.

Résumé

La conception d'un système de notation et de prédiction efficace pour les sports compétitifs attire de plus en plus d'attention. Une des approches consiste à inférer les forces des équipes à partir des résultats historiques des rencontres. Ceci est l'un des problèmes fondamentaux dans l'analyse sportive. Dans les années cinquante, Arpad Elo a proposé un algorithme simple pour la notation des joueurs d'échecs, qui est depuis l'algorithme de notation le plus populaire. Dans ce travail, nous expliquons le modèle mathématique derrière l'algorithme Elo, et en particulier, nous expliquons la supposition implicite, mais non mentionnée, du modèle avec égalité. De plus, nous proposons une extension du modèle, le rendant ainsi plus flexible, capable de prendre en considération d'autres suppositions plus réalistes. Ceci nous donne le nouvel algorithme, que nous appelons κ -Elo, qui garde la simplicité de l'algorithme Elo tout en ayant la possibilité de s'ajuster à la fréquence des égalités. Nous présentons une discussion sur l'importance du choix approprié des paramètres ainsi que des exemples illustreurs basés sur les résultats de la Première Ligue Anglaise de football et de la Ligue Nationale de Hockey.

Mots-clés Système de notation, Elo

Abstract

The design of an efficient rating and prediction system for competitive sports attracts a lot of attention. One of the approaches consists of inferring the strength of the teams from historic confrontation results. This is one of the fundamental problems in sports analytics. In the fifties, Arpad Elo proposed a simple, non-trivial rating algorithm for chess players, and since then, it has been the most popular rating algorithm. In this work, we explain the mathematical model behind the Elo algorithm, in particular we explain the implicit assumption yet not spelled out of the draw model. Furthermore, we propose an extension to the model, making it more flexible and thus capable of taking in consideration more realistic assumptions. This yield the new algorithm, we call κ -Elo, which is as simple as the Elo algorithm and it provides the possibility to adjust the draws frequency. We discuss the importance of fine-tuning the parameters and present illustrative examples from the English Premier League football games and from the National Hockey League games.

keywords : Rating system, Elo

Table des matières

Remerciements	iii
Résumé	v
Abstract	vii
Table des matières	ix
Liste des figures	xi
1 Introduction	1
1.1 Notation et prédiction	1
1.1.1 Concept et motivation	1
1.1.2 Comparaison par paires et l’algorithme Elo	2
1.2 Objectifs et problématique	3
1.3 Contribution	4
1.4 Structure du document	4
2 Principe de notation et l’algorithme Elo	5
2.1 Modèle gain-perte	6
2.2 Notation de maximum de vraisemblance	7
2.3 La méthode de gradient	9
2.4 Gradient stochastique et Elo	10
2.4.1 Gradient stochastique	11
2.4.2 L’algorithme Elo	12
3 Égalité	15
3.1 Explication de l’égalité dans l’algorithme Elo	16
3.2 Généralisation de l’algorithme Elo	20
3.2.1 Adaptation	21
3.2.2 κ dans κ -Elo	23
4 Exemple numérique	27
5 Conclusion et extensions possibles	35
Références	37

Liste des figures

2.1	La forme des fonctions F , F' et ξ du modèle de Thurstone et du modèle de Bradley-Terry, respectivement	10
4.1	La forme des fonctions de prédiction de κ -Elo pour différentes valeurs de η	28
4.2	Évolution des notes $\hat{\theta}_{m,n}$ de quelques équipes de la première ligue anglaise de football au cours de la saison 2013 – 2014	29
4.3	Score logarithmique de la saison 2013 – 2014 de la première ligue anglaise de football	31
4.4	Score logarithmique de la saison 2017 – 2018 de la première ligue anglaise de football	32
4.5	Évolution des notes $\hat{\theta}_{m,n}$ de quelques équipes de la ligue national de hockey au cours de la saison 2017 – 2018	33
4.6	Score logarithmique de la saison 2017 – 2018 de la ligue nationale de hockey	34

Chapitre 1

Introduction

1.1 Notation et prédiction

1.1.1 Concept et motivation

La prédiction des résultats des matchs dans les sports de compétition est d'une importance majeure et attire l'attention de tous les partis impliqués. Les bookmakers et les parieurs ont besoin des prédictions assez précises afin de maximiser leurs profits sachant les quantités colossales d'argent mises au jeu. Les entraîneurs et les propriétaires des équipes ont besoin des systèmes de notation et de prédiction pour les aider à surveiller et évaluer la performance de leurs équipes. De plus, les prédictions publiées dans les journaux sportifs et les réseaux sociaux avant le déroulement d'un match sont toujours un sujet de débats intéressant pour les supporters.

Le résultat d'une rencontre entre deux équipes dépend d'un grand nombre de facteurs. Prendre en considération tous les facteurs pertinents connaissant l'incertitude liée aux mesures est difficile et un bon système devra être capable d'apprendre à partir des observations historiques à caractère aléatoire. Le cadre probabiliste décrit comment représenter et manipuler l'incertitude reliée au modèle et à la prédiction. En effet, l'apprentissage des forces des équipes peut être considéré comme une inférence des modèles les plus plausibles pour expliquer les données, et un ordinateur pourra ensuite utiliser ces modèles pour faire des prédictions futures qui tiennent compte du caractère incertain des

données. Par conséquent, la notation et la prédiction des matchs des sports de compétition relèvent du domaine de la modélisation probabiliste.

La notation consiste à attribuer une valeur numérique (une *note*) à une équipe en utilisant les résultats des matchs passés. La notation et la prédiction sont étroitement connectées. La notation utilise les données (que nous supposons pertinentes) pour estimer les forces des équipes dans le but de les classer, tandis que la prédiction utilisera les forces inférées afin de prédire le résultat d'un match futur. Ces mesures des forces des équipes ne sont pas directement accessibles. Nous n'avons accès qu'aux résultats des rencontres et des statistiques reliées à ces dernières. Dans la majorité des sports compétitifs, telles que le football et le hockey, chaque résultat de match est le fruit de la confrontation d'une paire d'équipes, et toutes les équipes de la même compétition ne vont pas forcément s'affronter. Les modèles de comparaison par paires sont les plus adéquats pour analyser de telles données.

1.1.2 Comparaison par paires et l'algorithme Elo

Les modèles de comparaison par paires sont utilisés pour analyser les données des événements dans lesquelles la comparaison des compétiteurs n'est possible qu'en bloc de deux. Tel est le cas de la majorité des sports de compétition. L'objectif inférentiel de l'analyse de comparaison par paires est la notation et la prédiction d'un match futur. Les deux modèles comparaison par paires les plus populaires sont le modèle Thurstone (1927) et le modèle Bradley & Terry (1952), et ils sont utilisés dans la conception de la majorité des systèmes de notation.

Le système Elo (2008) (ou l'une de ses variantes) est le système de notation le plus populaire et le plus utilisé dans les sports compétitifs. La version originale proposée par Arpad Elo utilisait le modèle Thurstone (1927) pour estimer les probabilités des résultats. Ce dernier suppose que les forces des équipes dans chaque rencontre sont des variables aléatoires de distribution normale centrées autour de la vraie performance de l'équipe, et suppose que la variance est constante et est la même pour toutes les équipes. Une autre version du système qui utilise le modèle de comparaison par paires de Bradley & Terry (1952) remonte aux travaux de Zermelo (1929). L'algorithme Elo a bouleversé les techniques traditionnelles de notation dans les compétitions sportives qui se contentaient d'attribuer des points pour l'équipe gagnante (par exemple, dans les tournois de football l'équipe gagnante reçoit trois points et la perdante ne reçoit rien, et les deux équipes reçoivent un point en cas d'égalité),

Elo utilise non seulement le résultat du match en cours, mais aussi les notes des équipes avant la rencontre. Depuis (1970), le système Elo (plus spécifiquement la version qui utilise le modèle de Bradley & Terry (1952)) a été adopté par les fédérations d'échecs du monde entier. De plus, en 2018 l'algorithme Elo sous le nom "SUM" a été adopté par la Fédération Internationale de Football Association (FIFA) pour la notation des équipes nationales (FIFA (2019)) et est le système de notation le plus populaire pour beaucoup d'autres sports de compétition tels que le baseball, le football américain, le basket-ball... Vu son importance et sa popularité, l'algorithme Elo mérite une attention particulière, surtout parce qu'il est généralement présenté sans explication de ses fondements mathématiques.

1.2 Objectifs et problématique

L'objectif ultime de chaque système de notation est la prédiction la plus précise possible des résultats des matchs. Dans ce projet, nous adoptons le point de vue probabiliste pour la modélisation des résultats des matchs (la vraisemblance relie conditionnellement le résultat avec les notes correspondantes). Ce qui fait que nous pouvons utiliser les stratégies traditionnelles d'estimation (maximum de vraisemblance) pour trouver les notes. En ayant un modèle bien défini, la note obtenue peut être utilisée pour la prédiction.

Nous sommes particulièrement intéressés par la modélisation mathématique des égalités. Ce problème a été abordé en utilisant deux approches différentes: dans Rao & Kupper (1967) en utilisant un seuillage et dans Davidson (1970) en utilisant une approche axiomatique. Nous notons que les égalités ne sont pas modélisées dans l'algorithme Elo (2008). En effet, il n'y a même pas de modèle explicite des résultats, la dérivation de l'algorithme est le fruit de la forte intuition de l'auteur, sans qu'il ne présente une définition formelle d'un critère d'optimalité. Il a été découvert ultérieurement que la version courante de l'algorithme Elo (qui utilise le modèle de Bradley & Terry (1952)) trouve comme note une estimation du maximum de vraisemblance dans le cas d'un match à résultat binaire (gain, perte) (Király & Qian (2017)), par contre dans sa version originale l'algorithme Elo propose d'utiliser un modèle gaussien. Quant à l'égalité, l'algorithme Elo utilise le score fractionnel du match (par exemple l'égalité comme étant un demi-gain) Wikipedia (2019). Cependant, comme le modèle n'est pas explicitement spécifié, ceci constitue une faille logique: d'une part l'algorithme Elo traite

l'égalité, de l'autre part, il n'y a pas de modèle qui nous permet de calculer/prédire la probabilité d'égalité. L'objectif de cette mémoire est de combler cette lacune.

1.3 Contribution

Notre contribution principale consiste à définir le modèle mathématique ainsi que l'hypothèse implicite, mais non mentionnée derrière l'algorithme Elo dans le cadre des rencontres avec égalités. À partir de ce modèle, nous proposons une extension le rendant ainsi plus flexible, capable de prendre en considération d'autres hypothèses plus réalistes. Ceci nous donne le nouvel algorithme, que nous appelons κ -Elo, qui garde la simplicité de l'algorithme Elo tout en ayant la possibilité de s'ajuster à la fréquence des égalités. Nous nous proposons aussi de clarifier l'interprétation de l'algorithme Elo, ainsi qu'une discussion concernant le modèle proposé dans sa version originale (gaussienne).

1.4 Structure du document

La structure de notre mémoire est comme suit:

Chapitre 1	Mise en contexte du projet de mémoire ainsi qu'un aperçu sur la comparaison par paires et le système de notation le plus populaire Elo.
Chapitre 2	Nous présentons le modèle ainsi que la technique d'inférence que nous adoptons pour les matchs de type gain-perte. Nous présentons l'algorithme Elo et le modèle mathématique derrière sa dérivation ainsi qu'une discussion sur sa version originale gaussienne. Nous proposons une interprétation des résultats de l'algorithme Elo
Chapitre 3	Nous présentons le modèle mathématique derrière la dérivation de l'algorithme Elo pour les matchs avec égalité, ainsi qu'une formulation explicite de la prédiction des matchs futurs. Nous proposons le nouvel algorithme κ -Elo comme une généralisation de l'algorithme Elo qui est capable de s'ajuster à la fréquence des égalités entre les équipes.
Chapitre 4	Exemples numérique ainsi qu'une discussion à propos de κ -Elo.
Chapitre 4	Conclusion du mémoire.

Chapitre 2

Principe de notation et l’algorithme

Elo

Nous considérons M équipes qui s’affrontent une à une, chaque équipe est indexée par $m = 1, \dots, M$. À l’instant n nous observons le résultat y_n de la rencontre entre les deux équipes adverses défini par la paire $\mathbf{R}_n = \{i_n, j_n\}$. Nous considérons les trois résultats suivants:

- (i) l’équipe i_n gagne contre l’équipe j_n ; ce qu’on dénote par $\{i_n \succ j_n\}$ qui correspond à $y_n = 1$.
- (ii) l’équipe i_n égalise avec l’équipe j_n ; ce qu’on dénote par $\{i_n \doteq j_n\}$ qui correspond à $y_n = 0$.
- (iii) l’équipe i_n perd contre l’équipe j_n ; ce qu’on dénote par $\{i_n \prec j_n\}$ qui correspond à $y_n = -1$.

Pour simplifier la notation dans les dérivations futures, nous définissons les trois variables suivantes (qui indiquent respectivement le gain, égalité et perte)

$$g_n = \mathbb{1}[y_n = 1], \quad e_n = \mathbb{1}[y_n = 0], \quad p_n = \mathbb{1}[y_n = -1], \quad (2.1)$$

avec $\mathbb{1}[\cdot]$ la fonction indicatrice: $\mathbb{1}[C] = 1$ si la condition C est vraie sinon $\mathbb{1}[C] = 0$.

Après l’observation des résultats des matchs $y_k, k = 1, \dots, n$ nous voulons noter les équipes, ce qui revient à assigner un nombre réel θ_m à chacune. Cette note devra représenter les chances de gagner d’une équipe dans un match donné ; c’est pour cela qu’elle est considérée dans la littérature comme étant la «force» Glickman (1999) ou bien la «compétence» Caron & Doucet (2012). Ces notes θ_m devront être interprétées dans le cadre probabiliste. Autrement dit, les notes θ_m deviennent les

paramètres d'un système qui explique les résultats historiques (notation) et est capable de fournir une description probabiliste des résultats futurs (prédiction).

2.1 Modèle gain-perte

Dans cette section, nous commençons par formuler le modèle pour le cas binaire $y_n \in \{-1, 1\}$; et nous abordons la possibilité d'égalité dans le chapitre 3. La modélisation statistique consiste à définir la fonction de vraisemblance qui formule la probabilité d'avoir une observation donnée connaissant les paramètres du modèle. Autrement dit, nous voudrions établir la relation probabiliste reliant le résultat d'un match y_n (entre l'équipe i_n et j_n) aux notes correspondantes (respectivement θ_i et θ_j);

$$\Pr \{i_n > j_n\} = \bar{\Phi}_G(\theta_i, \theta_j), \quad \Pr \{i_n < j_n\} = \bar{\Phi}_P(\theta_i, \theta_j). \quad (2.2)$$

Afin de satisfaire la cohérence du système de notation, la fonction $\bar{\Phi}_G(\theta_i, \theta_j)$ doit satisfaire le postulat suivant : la croissance de la différence des notes entre l'équipe i et l'équipe j doit impérativement correspondre à une croissance de la probabilité de gain de l'équipe i contre l'équipe j ; ceci est satisfait si

$$\bar{\Phi}_G(\theta_i, \theta_j) = \Phi_G(\theta_i - \theta_j), \quad (2.3)$$

où $\Phi_G(v)$ est une fonction croissante bornée $0 \leq \Phi_G(v) \leq 1$, qui satisfait $\lim_{v \rightarrow +\infty} \Phi_G(v) = 1$ et $\lim_{v \rightarrow -\infty} \Phi_G(v) = 0$. Ces conditions sont satisfaites en adoptant $\Phi_G(v) = \Phi(v)$, où $\Phi(v)$ est une fonction de répartition adéquate de notre choix. Par symétrie $\Pr \{i_n > j_n\} = \Pr \{j_n < i_n\}$. La loi totale de probabilité impose que $\Pr \{i_n > j_n\} + \Pr \{i_n < j_n\} = 1$; ceci n'est satisfait que si la distribution de probabilité est paire $\Phi'(v) = \Phi'(-v)$ et nous obtenons

$$\Phi_G(v) = \Phi(v), \quad \Phi_P(v) = \Phi(-v) = 1 - \Phi(v). \quad (2.4)$$

Ces types de modèles ont été appelés modèles linéaires par David (1988). Un des choix populaires de la fonction $\Phi(v)$ est la fonction de répartition gaussienne Thurstone (1927)

$$F_G(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v/\sigma} e^{-\frac{t^2}{2}} dt, \quad (2.5)$$

ou bien une fonction de répartition de type logistique Bradley & Terry (1952)

$$F_L(v) = \frac{1}{1 + 10^{-v/\sigma}}, \quad (2.6)$$

où, dans les deux cas, σ est un facteur d'échelle.

En reprenant les fonctions indicatrices de (2.1), la formulation finale du modèle est comme suit:

$$\Pr \{Y_n = y_n\} = g_n \Phi(\theta_{i_n} - \theta_{j_n}) + p_n \Phi(\theta_{j_n} - \theta_{i_n}), \quad (2.7)$$

et le modèle ne dépend que de la différence des notes, θ_{i_n} et θ_{j_n} ce qui fait que la notation est arbitraire selon les paramètres suivants ;

- (i) le choix de l'échelle, dans le sens qu'une note θ_m obtenue avec un facteur d'échelle σ peut être transformée en une note θ'_m avec un autre facteur d'échelle σ' selon la multiplication: $\theta'_m = \theta_m \sigma' / \sigma$.
- (ii) le choix de l'origine, puisque la différence des notes $\theta_{i_n} - \theta_{j_n}$ de l'équation (2.4) est invariante par rapport à l'addition d'un nombre réel quelconque θ_0 à chaque élément.
- (iii) la base de l'exposant de l'équation (2.6) ; par exemple, $10^{-v/\sigma} = e^{-v/\sigma^*}$ avec $\sigma' = \sigma \log_{10} e$, ce qui fait que le changement de la base 10 à la base népérienne ne nécessite que le changement du facteur d'échelle σ en σ' .

Après la définition du problème, il faut choisir une stratégie d'inférence afin de résoudre le problème en question. Il existe différentes stratégies d'inférence. Nous avons opté pour l'utilisation du principe de maximum de vraisemblance.

2.2 Notation de maximum de vraisemblance

En utilisant les résultats obtenus dans la section 2.1, nous reformulons le modèle reliant la variable aléatoire Y_n avec les notes de la façon suivante:

$$\Pr(Y_n = y_n | \mathbf{x}_n, \boldsymbol{\theta}) = g_n \Phi(v_n) + p_n \Phi(-v_n), \quad (2.8)$$

$$v_n = v_n(\boldsymbol{\theta}) = \mathbf{x}_n^T \boldsymbol{\theta} = \theta_{i_n} - \theta_{j_n}, \quad (2.9)$$

où $\boldsymbol{\theta} = [\theta_1, \dots, \theta_M]^T$ est le vecteur de toutes les notes, $(\cdot)^T$ est la transposée, et \mathbf{x}_n est le vecteur de planification des matchs,

$$\mathbf{x}_n = [0, \dots, \underbrace{1}_{i_n\text{-ième pos.}}, 0, \dots, 0, \underbrace{-1}_{j_n\text{-ième pos.}}, 0, \dots, 0]^T. \quad (2.10)$$

Notre objectif maintenant est de trouver la valeur de la note $\boldsymbol{\theta}$ à l'instant n sachant les résultats des matchs passés $\{y_l\}_{l=1}^n$ et le vecteur de planification correspondant \mathbf{x}_n . Ceci est principalement un problème d'estimation de paramètres. L'estimation de maximum de vraisemblance de $\boldsymbol{\theta}$ à l'instant n est obtenue en solvant le problème d'optimisation suivant

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \mathbf{J}_n(\boldsymbol{\theta}), \quad (2.11)$$

où $\hat{\boldsymbol{\theta}}_n = [\hat{\boldsymbol{\theta}}_{n,1}, \dots, \hat{\boldsymbol{\theta}}_{n,M}]^T$, et

$$\mathbf{J}_n(\boldsymbol{\theta}) = -\log \Pr \left\{ \{Y_l\}_{l=1}^n = \{y_l\}_{l=1}^n \mid \boldsymbol{\theta}, \{\mathbf{x}_l\}_{l=1}^n \right\}. \quad (2.12)$$

En supposant que les résultats des matchs Y_l sont conditionnellement indépendants sachant les notes $\boldsymbol{\theta}$, $\Pr \left\{ \{Y_l\}_{l=1}^n = \{y_l\}_{l=1}^n \mid \boldsymbol{\theta}, \{\mathbf{x}_l\}_{l=1}^n \right\} = \prod_{l=1}^n \Pr \{Y_n = y_n \mid \boldsymbol{\theta}, \mathbf{x}_n\}$, le problème d'optimisation devient

$$\mathbf{J}_n(\boldsymbol{\theta}) = \sum_{l=1}^n \mathbf{L}_l(\boldsymbol{\theta}), \quad (2.13)$$

avec

$$\mathbf{L}_l(\boldsymbol{\theta}) = -\log \Pr \{Y_l = y_l \mid \boldsymbol{\theta}\} = -g_l \log \Phi(\mathbf{x}_l^T \boldsymbol{\theta}) - p_l \log \Phi(-\mathbf{x}_l^T \boldsymbol{\theta}), \quad (2.14)$$

où nous avons appliqué le modèle (2.8).

2.3 La méthode de gradient

Puisque $J_n(\boldsymbol{\theta})$ est une fonction convexe Bagnoli & Bergstrom (2005), le problème (2.11) admet un minimum global qui peut être atteint en utilisant l'algorithme du gradient descendant qui donne la mise à jour du paramètre $\hat{\boldsymbol{\theta}}_n$ comme suit:

$$\hat{\boldsymbol{\theta}}_n \leftarrow \hat{\boldsymbol{\theta}}_n - \mu \nabla_{\boldsymbol{\theta}} J_n(\hat{\boldsymbol{\theta}}_n), \quad (2.15)$$

que nous itérons jusqu'à la convergence pour un n donné ; le gradient est calculé de la façon suivante

$$\nabla_{\boldsymbol{\theta}} J_n(\boldsymbol{\theta}) = \sum_{l=1}^n \nabla_{\boldsymbol{\theta}} L_l(\boldsymbol{\theta}), \quad (2.16)$$

et le pas μ devra être choisi convenablement afin de garantir la convergence. En utilisant (2.14) nous obtenons

$$\nabla_{\boldsymbol{\theta}} L_l(\boldsymbol{\theta}) = -g_l \mathbf{x}_l \psi(\mathbf{x}_l^T \boldsymbol{\theta}) + p_l \mathbf{x}_l \psi(-\mathbf{x}_l^T \boldsymbol{\theta}) = -\mathbf{x}_l \epsilon_l(v_l) \quad (2.17)$$

avec

$$\psi(v) = \frac{d}{dv} \log \Phi(v) = \frac{\Phi'(v)}{\Phi(v)}, \quad (2.18)$$

et

$$\epsilon_l(v_l) = g_l \psi(v_l) + (g_l - 1) \psi(-v_l) = [g_l - \Phi(v_l)] \xi(v_l), \quad (2.19)$$

où

$$\xi(v) = \frac{\psi(v)}{\Phi(-v)} = \frac{\Phi'(v)}{\Phi(v) \Phi(-v)}. \quad (2.20)$$

En utilisant la fonction de répartition logistique (selon Bradley & Terry (1952)) (2.6) dans (2.20) nous obtenons:

$$\xi_L(v) = \frac{1}{\sigma'}, \quad (2.21)$$

et (2.19) devient

$$\epsilon_l(v_l) = \frac{1}{\sigma'} [g_l - \Phi(v_l)] \quad (2.22)$$

avec $\sigma' = \sigma \log_{10} e$.

En ce qui concerne la fonction de répartition gaussienne, la fonction $\xi_G(v)$ dans (2.20) n'a pas de solution analytique, mais elle peut être facilement calculée de façon numérique. La figure 2.1 montre que les fonctions $\xi_L(v)$ et $\xi_G(v)$ sont remarquablement différentes, même si, en effectuant une mise à l'échelle appropriée, la différence entre les fonctions de répartition F_G et F_L et la différence entre leurs dérivées F'_G et F'_L est presque négligeable.

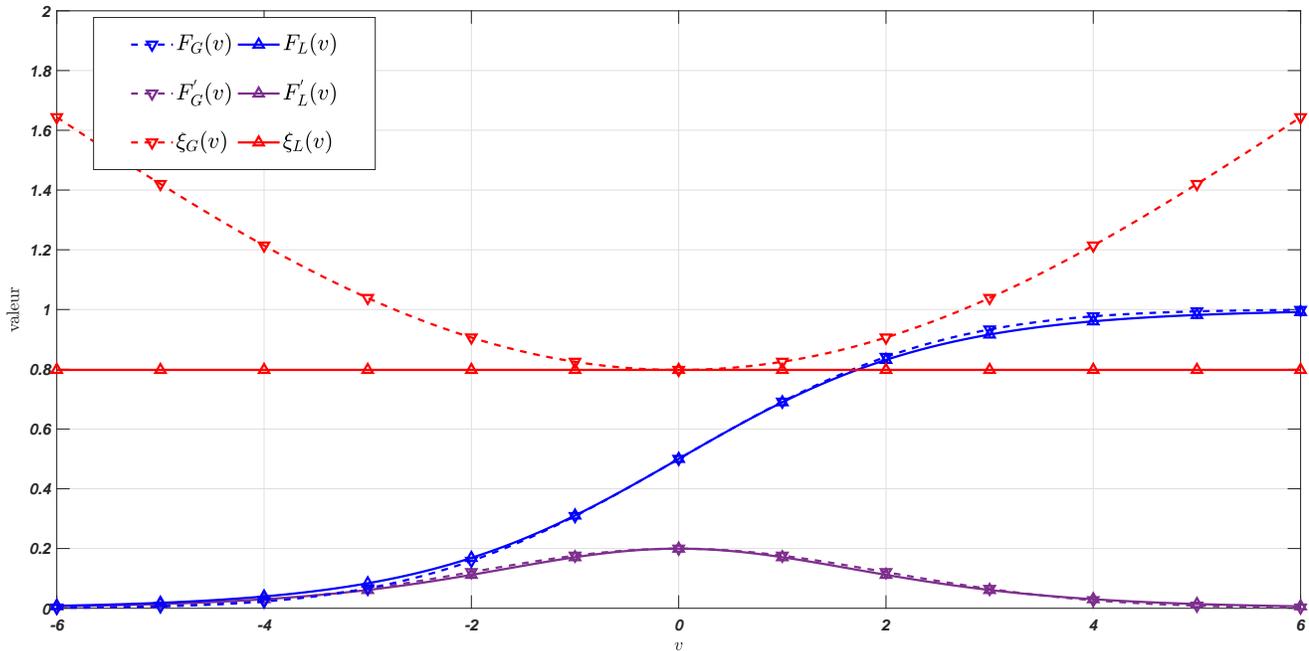


Figure 2.1 – La forme des fonctions F , F' et ξ du modèle de Thurstone et du modèle de Bradley-Terry, respectivement, avec un facteur d'échelle de la fonction de répartition gaussienne $\sigma_G = 2$, et le facteur d'échelle de la fonction de répartition logistique $\sigma_L = \sigma_G \sqrt{\pi/8 \log_e 10}$, valeur choisie de telle sorte que $F'_G(0) = F'_L(0)$

2.4 Gradient stochastique et Elo

La solution obtenue en (2.15) est développée à partir du modèle (2.8) qui nécessite que les notes θ restent constantes au cours des temps $l = 1, \dots, n$. Par contre, les notes des joueurs sont variables (amélioration de la force due à l'entraînement, des blessures, une nouvelle stratégie de l'entraîneur, etc). Un système de notation réaliste devra être capable de suivre la variation de la valeur des notes θ .

2.4.1 Gradient stochastique

Afin de suivre l'évolution des notes des équipes, la façon la plus simple serait d'utiliser le gradient stochastique descendant **SG** qui diffère du gradient descendant régulier dans les points suivants:

- (i) à l'instant n une seule itération du gradient descendant est exécutée.
- (ii) le gradient n'est calculé que pour le terme de l'observation en cours $L_n(\hat{\theta}_n)$.
- (iii) l'estimation disponible à l'instant n , $\hat{\theta}_n$ est utilisée comme point initial de la mise à jour suivante $n + 1$.

La formule de mise à jour du gradient stochastique **SG** se fait comme suit:

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \mu \nabla_{\theta} L_n(\theta) = \hat{\theta}_n + \mu \mathbf{x}_n \epsilon_n(v_n) \quad (2.23)$$

$$= \hat{\theta}_n + \mu \mathbf{x}_n [g_n - \Phi(v_n)] \xi(v_n), \quad (2.24)$$

où μ est le pas représentatif du taux d'apprentissage.

L'algorithme du gradient stochastique **SG** est très populaire et a été utilisé d'une manière extensive dans des domaines divers: dans les problèmes d'estimation de paramètres (Toulis & Airoidi (2014)), dans les implémentations des applications d'ingénieries où il est connu sous le nom de méthode des moindres carrés Farhang-Boroujeny (1998), et dans l'entraînement des réseaux de neurones sous le nom de l'algorithme de rétropropagation du gradient Rumelhart *et al.* (2012).

À un instant donné, chaque équipe joue au plus un match, donc le vecteur de planification \mathbf{x}_n est creux et n'a que deux termes non nuls, et à chaque itération, seules les notes de l'équipe i_n et l'équipe j_n seront modifiées. Ce qui fait que la formule de mise à jour du vecteur des notes de toutes les équipes $i \in \{i_n, j_n\}$ (2.24) peut être simplifiée de la manière suivante:

$$\hat{\theta}_{n+1,i} = \hat{\theta}_{n,i} + \mu [s_i - \Phi(\Delta_i)] \xi(\Delta_i) \quad (2.25)$$

avec $\Delta_i = \hat{\theta}_{n,i} - \hat{\theta}_{n,j}$ et j est l'index de l'équipe adverse à i , et $s_i = \mathbb{1}[i > j]$ une variable indicatrice de l'équipe gagnante ($s_i = 1$ si l'équipe i gagne et $s_i = 0$ si elle perd).

Nous observons ce qui suit:

- (i) Puisque la somme des éléments de \mathbf{x}_n est nulle ($\sum_m \mathbf{x}_m = \mathbf{1}\mathbf{x}_n = 0$, où $\mathbf{1} = [1, \dots, 1]$), la somme des notes $\boldsymbol{\theta}$ ne change pas au cours du temps,

$$\mathbf{1}\hat{\boldsymbol{\theta}}_{n+1} = \mathbf{1}\hat{\boldsymbol{\theta}}_n + \mathbf{1}\mathbf{x}_n\mu[g_n - \Phi(v_n)]\xi(v_n) = \mathbf{1}\hat{\boldsymbol{\theta}}_n. \quad (2.26)$$

- (ii) Le taux de variabilité de la note $\theta_{n,i}$ est affecté par la valeur du taux d'apprentissage, μ mais aussi par le terme multiplicatif $\xi(\Delta_i)$ dans (2.25). Nous constatons que le terme $[s_i - \Phi(\Delta_i)] \in (-1, 1)$ est borné en valeur absolue, et la différence entre les fonctions F' peut être négligeable (voir la figure 2.1). En utilisant une fonction de répartition logistique, $\Phi(v) = F_L(v)$ nous obtenons une fonction $\xi_L(v)$ de (2.21) constante. Par contre, la fonction $\xi(v)$ de (2.25) peut être très grande. La figure 2.1 montre que la fonction $\xi_G(v)$ qui correspond à la fonction de répartition gaussienne $\Phi(v) = F_G(v)$ est assez large et les notes correspondantes peuvent être arbitraires.

2.4.2 L'algorithme Elo

Dans la littérature, l'équation de mise à jour de l'algorithme Elo pour les jeux binaires est présentée comme suit:

$$\hat{\theta}_{n+1,i} = \hat{\theta}_{n,i} + K[s_i - F_{\text{Elo}}(\Delta_i)] \quad (2.27)$$

où le terme Δ_i est la même que dans (2.25); les seules différences avec (2.25) sont

- (i) le terme $\xi(\Delta_i)$ est éliminé.
- (ii) pour la compatibilité avec la littérature nous utilisons K à la place de μ
- (iii) $F_{\text{Elo}}(v)$, à la place de $\Phi(v)$, qui est le score moyen du match avant son déroulement (le match prédit).

Nous en tirons ainsi les deux conclusions suivantes:

1. En utilisant $\Phi(v) = F_L(v)$ et en appliquant l'algorithme du gradient stochastique **SG** (2.25), nous obtenons l'algorithme Elo (2.27) avec $F_{\text{Elo}}(v) = F_L(v)$ et $K = \mu\sigma^*$ (la valeur du facteur d'échelle σ' n'est pas pertinente puisque la valeur du pas est généralement choisie d'une manière heuristique). Ce qui fait que la version d'Elo la plus populaire, celle qui est basée sur $F_{\text{Elo}}(v) = F_L(v)$, implémente l'algorithme **SG** qui fournit un estimé du maximum de

vraisemblance des notes des équipes. Ceci a été observé précédemment par Király & Qian (2017).

2. En utilisant $\Phi(v) = F_G(v)$ nous obtenons la version suivante de (2.25)

$$\hat{\theta}_{n+1,i} = \hat{\theta}_{n,i} + \mu[s_i - F_G(\Delta_i)] \xi_G(\Delta_i) \quad (2.28)$$

qui est différente de l'algorithme Elo (2.27) avec $F_{Elo}(v) = F_G(v)$ puisque le terme $\xi_G(v)$ est manquant (comme il est variable, il ne peut pas être absorbé par le pas K). Donc la variante gaussienne de l'algorithme Elo n'implémente pas le gradient stochastique pour avoir une estimation du maximum de vraisemblance des notes des équipes. Choisir une fonction de répartition arbitraire à la place de $F_{Elo}(v)$ n'est pas méthodique. Il faut d'abord expliciter le modèle ensuite choisir une méthode d'inférence convenable. Il est vrai que la version la plus populaire d'Elo utilise une fonction logistique comme modèle, mais nous trouvons que c'est important de faire cette clarification vu que cela montre la signifiante de l'explicitation du modèle utilisé.

La deuxième conclusion nous pousse à nous poser la question suivante: quel est le problème que l'algorithme Elo résout quand une fonction arbitraire $F_{Elo}(v)$ est utilisée dans (2.27)? Et particulièrement, qu'est ce qui se passe si nous utilisons $F_{Elo}(v) = F_G(v)$ dans (2.27) comme ça était initialement proposé par le concepteur de l'algorithme Elo?

Pour obtenir la réponse à nos interrogations, nous revenons au problème résolu par le gradient stochastique et nous réécrivons le gradient (2.16) de tous les matchs en utilisant (2.17) de la manière suivante:

$$\nabla_{\theta} J(\theta) = \sum_{(i,j) \in I} m_{i,j} \mathbf{x}_{i,j} \left[p_{i,j}(1 - \Phi(\mathbf{x}_{i,j}^T \theta)) + (p_{i,j} - 1)\Phi(\mathbf{x}_{i,j}^T \theta) \right] \xi(\mathbf{x}_{i,j}^T \theta) \quad (2.29)$$

$$= \sum_{(i,j) \in I} m_{i,j} \mathbf{x}_{i,j} (p_{i,j} - \Phi(\mathbf{x}_{i,j}^T \theta)) \xi(\mathbf{x}_{i,j}^T \theta), \quad (2.30)$$

où $\mathbf{x}_{i,j}$ est le vecteur de planification correspondant à une rencontre entre l'équipe i et l'équipe j , $m_{i,j}$ la proportion des matchs qui ont comme vecteur de planification $\mathbf{x}_l = \mathbf{x}_{i,j}$, et $p_{i,j}$ est la proportion des matchs gagnés par l'équipe i contre l'équipe j (par exemple si les équipes i et j ont joué $m_{i,j} = 4$ matchs, et l'équipe i en a gagné trois alors $p_{i,j} = 0.75$). I est l'espace qui dénote toutes les paires (i, j) .

L'application du gradient (2.30) d'une manière itérative comme dans (2.15) produit la solution du maximum de vraisemblance. Par contre, le gradient descendant (2.15) en utilisant le gradient défini par l'algorithme Elo qui ignore le terme $\xi(\cdot)$ est:

$$\mathbf{g}_{\text{Elo}} = \sum_{(i,j) \in I} \mathbf{x}_{i,j} (p_{i,j} - F_{\text{Elo}}(\mathbf{x}_{i,j}^T \boldsymbol{\theta})), \quad (2.31)$$

avec \mathbf{g}_{Elo} le gradient d'une fonction inconnue. Nous pouvons maintenant observer que si $p_{i,j} = F_{\text{Elo}}(\mathbf{x}_{i,j}^T \boldsymbol{\theta}) = F_G(\mathbf{x}_{i,j}^T \boldsymbol{\theta}) = \Phi(\mathbf{x}_{i,j}^T \boldsymbol{\theta})$, le gradient de l'équation (2.30) et le gradient de l'algorithme Elo (2.31) s'annulent. Donc si le modèle gaussien est capable de représenter les données, l'algorithme Elo fournira la solution désirée. Par contre, l'algorithme Elo fournit généralement une approximation de l'estimateur de maximum de vraisemblance.

Dorénavant, quand nous parlons de l'algorithme Elo nous assumons que c'est le modèle logistique qui est utilisé.

Chapitre 3

Égalité

Dans ce chapitre, nous allons traiter des modèles avec des résultats ternaires en incluant l'égalité entre les deux équipes adverses. Bien que nous l'avions ignoré dans la modélisation précédente, l'égalité entre deux équipes est un résultat informatif qui doit être pris en considération, surtout par les systèmes de notation dédiés aux sports de compétition là où l'égalité est assez fréquente (tel que le football). Nous avons remarqué que ce problème a été contourné dans la littérature en utilisant des approches divers ; soit en ignorant l'égalité Langville & Meyer (2012), ce qui fait qu'il y a une sous-exploitation des données disponibles, ou bien en la considérant comme un gain partiel en adoptant un score fractionnel $s_i = \frac{1}{2}$ Langville & Meyer (2012) Glickman & Hennessy (2015). Bien que l'approche heuristique du gain partiel peut être utile pour l'inférence des notes, elle fait abstraction du modèle utilisé et le calcul de la prédiction à partir des notes inférées n'est pas possible.

Ainsi, nous avons opté pour la prise en compte explicite de l'égalité entre les deux équipes par le modèle; nous augmentons le modèle binaire en y incluant la probabilité conditionnelle de l'égalité.

$$\Pr \{i \doteq j | \theta_i, \theta_j\} = \Phi_E(\theta_i, \theta_j). \quad (3.1)$$

Afin de satisfaire la cohérence du système de notation, la fonction $\Phi_E(v)$ doit satisfaire le postulat suivant: la croissance de la différence des notes en valeur absolue entre l'équipe i et l'équipe j doit correspondre à la croissance de la probabilité de gain ou de perte ($\Phi_E(|v|)$ est décroissante), tandis

que la proximité des notes, $\theta_i \approx \theta_j$, doit correspondre à une probabilité d'égalité maximale ($\Phi_E(v)$ est maximale pour $v = 0$).

La loi totale de probabilité impose que

$$\Phi_G(v) + \Phi_P(v) + \Phi_E(v) = 1, \quad (3.2)$$

ce qui implique que les fonctions $\Phi_G(v)$ et $\Phi_P(v)$ sont différentes de celles utilisées dans le cas binaire et doivent être changées.

3.1 Explication de l'égalité dans l'algorithme Elo

L'algorithme Elo utilise l'approche du gain partiel en considérant un score fractionnel $s_i = \frac{1}{2}$ Langville & Meyer (2012) Elo (2008) sans définir la fonction $\Phi_E(v)$. C'est une situation confuse: d'un côté, l'algorithme Elo prend en considération les égalités, de l'autre côté, il n'y a pas de modèle qui nous permettrait de calculer la probabilité d'égalité à partir des notes inférées θ .

L'objectif de cette section est la rétro-ingénierie de l'algorithme Elo en trouvant un modèle gain/perte/égalité compatible avec son opération. Ceci nous permettrait de comprendre les résultats obtenus et de calculer les prédictions des matchs futurs.

Nous commençons par mettre au carré la loi totale de probabilité du cadre binaire traité précédemment $\Phi(v) + \Phi(-v) = 1$, nous obtenons:

$$\Phi^2(v) + \Phi^2(-v) + 2\Phi(v)\Phi(-v) = 1 \quad (3.3)$$

et donc, en adoptant

$$\Phi_G(v) = \Phi^2(v) \quad (3.4)$$

$$\Phi_P(v) = \Phi^2(-v) \quad (3.5)$$

$$\Phi_E(v) = 2\Phi(v)\Phi(-v), \quad (3.6)$$

nous satisfaisons la loi totale de probabilité (3.2) tout en gardant la symétrie du cas binaire $\Phi_G(v) = \Phi_P(-v)$.

Maintenant nous récrivons (2.14) en reflétant (3.4)-(3.6) comme suit

$$L_l(\boldsymbol{\theta}) = -\log \Pr \{Y_l = y_l | \boldsymbol{\theta}\} \quad (3.7)$$

$$= -g_l \log \Phi_G(v_l) - p_l \log \Phi_P(v_l) - e_l \log \Phi_E(v_l) \quad (3.8)$$

$$= -2g_l \log \Phi(v_l) - 2p_l \log \Phi(-v_l) - e_l \left[\log \Phi(v_l) + \log \Phi(-v_l) + \log(2) \right], \quad (3.9)$$

et le gradient est calculé comme dans (2.17)

$$\nabla_{\boldsymbol{\theta}} L_l(\boldsymbol{\theta}) = -2\tilde{g}_l \mathbf{x}_l \psi(v_l) + 2\tilde{p}_l \mathbf{x}_l \psi(-v_l) \quad (3.10)$$

$$= -2\mathbf{x}_l \tilde{\epsilon}_l(\boldsymbol{\theta}), \quad (3.11)$$

où $\tilde{g}_l = g_l + e_l/2$, $\tilde{p}_l = p_l + e_l/2$ et

$$\tilde{\epsilon}_l(\boldsymbol{\theta}) = \tilde{g}_l - \Phi(v_l). \quad (3.12)$$

Nous obtenons ainsi la même équation de mise à jour du cas binaire (2.24)

$$\hat{\boldsymbol{\theta}}_{n+1} = \hat{\boldsymbol{\theta}}_n + 2\mu \mathbf{x}_n [\tilde{g}_n - \Phi(v_n)] \quad (3.13)$$

$$= \hat{\boldsymbol{\theta}}_n - 2\mu \mathbf{x}_n [\tilde{p}_n - \Phi(v_n)], \quad (3.14)$$

qui est la même équation de mise à jour de l'algorithme Elo (2.27)

$$\hat{\boldsymbol{\theta}}_{n+1,i} = \hat{\boldsymbol{\theta}}_{n,i} + K[s_i - \Phi(\Delta_i)], \quad (3.15)$$

avec la nouvelle définition du score $s_i = 1$ pour l'équipe gagnante, $s_i = 0$ pour l'équipe perdante et $s_i = 0.5$ en cas d'égalité, et pour des raisons de compatibilité, la multiplication par 2 est absorbée par le pas K (la seule différence entre l'équation (3.12) et (2.22)).

Donc l'algorithme Elo implémente le gradient stochastique pour inférer les notes $\boldsymbol{\theta}$ en utilisant un estimé de maximum de vraisemblance sur le modèle défini dans (3.4)-(3.6).

Nous dérivons ainsi les deux observations suivantes:

1. Alors que la définition explicite du modèle ne change pas l'opération effectuée par l'algorithme, elle nous permet de clarifier l'interprétation de ces résultats. À savoir, connaissant les notes $\hat{\theta}_n$ la prédiction d'un match futur devra être estimée de la manière suivante:

$$\Pr \{i > j | \hat{\theta}_i, \hat{\theta}_j\} = \Phi^2(\hat{\theta}_i - \hat{\theta}_j) \quad (3.16)$$

$$\Pr \{i < j | \hat{\theta}_i, \hat{\theta}_j\} = \Phi^2(\hat{\theta}_j - \hat{\theta}_i) \quad (3.17)$$

$$\Pr \{i \doteq j | \hat{\theta}_i, \hat{\theta}_j\} = 2 \Phi(\hat{\theta}_i - \hat{\theta}_j) \Phi(\hat{\theta}_j - \hat{\theta}_i). \quad (3.18)$$

2. Nous soulignons que s_i est l'indicateur du résultat du match, en utilisant (3.4)-(3.6) nous pouvons calculer sa valeur moyenne

$$\mathbb{E}_{Y_l | \hat{\theta}_{l,i}, \hat{\theta}_{l,j}} [s_i(Y_l)] = \Pr \{i > j | \hat{\theta}_i, \hat{\theta}_j\} + \frac{1}{2} \Pr \{i \doteq j | \hat{\theta}_i, \hat{\theta}_j\} \quad (3.19)$$

$$= [\Phi(\Delta_i)]^2 + \Phi(\Delta_i)\Phi(-\Delta_i) \quad (3.20)$$

$$= \Phi(\Delta_i)[\Phi(\Delta_i) + \Phi(-\Delta_i)] = \Phi(\Delta_i). \quad (3.21)$$

Donc effectivement, la fonction $\Phi(\Delta_i)$ de la mise à jour de l'algorithme Elo (2.27) est la valeur moyenne du score s_i . À notre connaissance, la dérivation mathématique de l'interprétation de la fonction $\Phi(\Delta_i)$ n'a pas été présentée dans la littérature, et ceci est dû au traitement implicite des égalités. Cependant, le concepteur de l'algorithme Elo a fait preuve d'une intuition formidable en définissant correctement les termes de son algorithme sans faire référence au modèle probabiliste correspondant. Nous soulignons encore que la notion de score moyen s_i n'est pas nécessaire dans le développement du gradient stochastique, et le fait que la valeur du score fractionnel $s_i = \frac{1}{2}$ est due à la forme particulière de la probabilité conditionnelle (3.6) et l'absorption du pas K de la multiplication par 2 (3.11).

Alors que la clarification de l'interprétation du score est utile, la première observation concernant l'interprétation explicite du résultat de l'algorithme est la plus importante. Nous rappelons que dans le cas binaire, la fonction $\Phi(\Delta_i)$ correspondait à la probabilité de gain d'un match (2.4). Par contre, dans le cas ternaire, une telle interprétation est incorrecte puisque la probabilité de gain d'un match est donnée par (3.4). Nous allons voir dans le chapitre 4 que l'utilisation de la fonction $\Phi(\Delta_i)$ comme étant la probabilité de gain donne de piètres résultats. Par contre, il est possible de reformuler l'algorithme Elo afin de faire apparaître la probabilité de gain. En utilisant la loi totale

de probabilité (3.3) et le fait que $\Phi(-v) = 1 - \Phi(v)$ nous avons

$$\Phi(v) = \frac{1 + \Phi^2(v) - \Phi^2(-v)}{2}, \quad (3.22)$$

et donc l'équation de mise à jour de l'algorithme Elo (3.15) peut être reformulée de la façon suivante (en sortant la multiplication par 2 précédemment absorbée par K)

$$\hat{\theta}_{n+1,i} = \hat{\theta}_{n,i} + K[2s_i - 1 + [\Phi^2(-\Delta_i) - \Phi^2(\Delta_i)]] \quad (3.23)$$

$$= \hat{\theta}_{n,i} + K[y_i + [\Pr \{i < j | \hat{\theta}_i, \hat{\theta}_j\} - \Pr \{i > j | \hat{\theta}_i, \hat{\theta}_j\}]] \quad (3.24)$$

avec $y_i \in \{1, 0, -1\}$ une variable indicatrice du résultat de la rencontre de l'équipe i avec l'équipe j (respectivement gain, égalité et perte). Donc la mise à jour de l'algorithme Elo est une incrémentation de la note par la différence de la prédiction de gain et de la perte (translatée par y_i) multipliée par le pas K .

Nous avons remarqué que la confusion par rapport à la prédiction de l'égalité est persistante dans la littérature, et ceci est probablement dû à la modélisation implicite de l'algorithme Elo (que nous présentons (3.4)-(3.6)) et sa dérivation explicite où Elo (2008) ne considère pas l'égalité dans le cadre probabiliste formel. Cette difficulté conceptuelle a été observée par d'autres travaux comme Glickman (1999) ou Lasek *et al.* (2013). En particulier, (Glickman, 1999) utilise une fonction $\Phi_E(v) = \sqrt{\Phi(v)\Phi(-v)}$ tout en gardant l'héritage de l'algorithme Elo binaire ($\Phi_G(v) = \Phi(v)$ et $\Phi_P(v) = \Phi(-v)$), ce qui donne une solution approximative puisque la loi totale de probabilité (3.2) n'est pas satisfaite.

Nous retirons la leçon suivante: malgré la simplicité apparente de l'algorithme Elo, nous ne pouvons pas modifier ses paramètres sans la formulation du modèle correspondant. Bien que nous ayons expliqué le choix de la valeur du score fractionnel $s_i = \frac{1}{2}$, rien ne garantit que la modification arbitraire du score s_i correspondra à un modèle probabiliste particulier. Par conséquent, au lieu de modifier les paramètres de l'algorithme Elo (2.27), la modification devra être faite à partir du modèle lui-même.

3.2 Généralisation de l'algorithme Elo

Après la modélisation implicite de l'égalité dans l'algorithme Elo, un nouveau problème apparaît directement. À savoir, en prenant en considération l'égalité, nous avons trois événements (donc deux probabilités à estimer) mais l'algorithme Elo n'a que deux degrés de liberté et n'est pas capable de bien représenter les données ternaires. Par exemple, en utilisant (3.16)-(3.18) la prédiction d'une rencontre entre deux équipes à note égale $\hat{\theta}_i = \hat{\theta}_j$, la prédiction sera toujours $\Pr \{i \succ j | \hat{\theta}_i, \hat{\theta}_j\} = 0.25$ et $\Pr \{i \doteq j | \hat{\theta}_i, \hat{\theta}_j\} = 0.5$. L'algorithme Elo fait cette supposition d'une manière implicite, mais il n'y a aucune raison qui nous pousse à adopter cette solution rigide qui pourra être inadéquate pour la représentation des données. Une approche plus générale est nécessaire.

Une alternative a été proposée par Rao & Kupper (1967) et a été utilisée ultérieurement dans Király & Qian (2017), Herbrich & Graepel (2006) et Fahrmeir & Tutz (2006), ils modifient le modèle en utilisant une variable de seuil v_0 de la manière suivante:

$$\Phi_G(v) = \Phi(v - v_0), \quad \Phi_P(v) = \Phi_G(-v), \quad \Phi_E(v) = \Phi(v + v_0) - \Phi(v - v_0). \quad (3.25)$$

Bien que (3.25) est définitivement utile et résout le problème de l'égalité d'une manière formelle tout en ayant trois degrés de liberté, nous estimons qu'il n'est pas une généralisation adéquate de l'algorithme Elo puisqu'il n'existe pas un paramètre v_0 qui transforme (3.25) en (3.4)-(3.6) (le modèle derrière l'algorithme Elo).

Ainsi nous proposons plutôt d'utiliser le modèle de Davidson (1970) qui peut être défini de la manière suivante:

$$\Phi_G(v) = \Phi_\kappa(v) = \frac{10^{0.5v/\sigma}}{10^{0.5v/\sigma} + 10^{-0.5v/\sigma} + \kappa} \quad (3.26)$$

$$\Phi_P(v) = \Phi_\kappa(-v) = \frac{10^{-0.5v/\sigma}}{10^{0.5v/\sigma} + 10^{-0.5v/\sigma} + \kappa} \quad (3.27)$$

$$\Phi_E(v) = \kappa \sqrt{\Phi_G(v)\Phi_P(v)} = \frac{\kappa}{10^{0.5v/\sigma} + 10^{-0.5v/\sigma} + \kappa}, \quad (3.28)$$

où $\kappa \geq 0$ est un paramètre d'adaptation de l'égalité (le troisième degré de liberté).

Nous soulignons que le modèle (3.26)- (3.28) n'est pas nécessairement meilleur que le modèle (3.25) dans sa capacité à représenter les données. Notre motivation derrière l'adoption du modèle

(3.26)- (3.28) est qu'il généralise le modèle derrière l'algorithme Elo. En effet, pour $\kappa = 0$ nous obtenons le modèle correspondant à la dérivation de l'algorithme Elo binaire (2.27), et pour $\kappa = 2$ nous avons

$$\Phi_G(v) = \frac{10^{0.5v/\sigma}}{\left(10^{0.25v/\sigma} + 10^{-0.25v/\sigma}\right)^2} = \Phi^2(v/2) \quad (3.29)$$

qui à un facteur près σ , correspond au modèle ternaire implicite de l'algorithme Elo (3.4)-(3.6).

Donc, la modélisation implicite derrière l'algorithme Elo ternaire correspond à un cas particulier de la modélisation explicite de l'égalité de Davidson (1970) ($\kappa = 2$).

3.2.1 Adaptation

Nous notons que la fonction $-\log \Phi_\kappa(v)$ est convexe Davidson (1970), donc le gradient basé sur l'adaptation est convergent pour un pas μ adéquat.

Pour la dérivation du nouvel algorithme d'adaptation, nous recalculons (3.7)

$$L_l(\boldsymbol{\theta}) = -\log \Pr \{Y_l = y_l | \boldsymbol{\theta}\} \quad (3.30)$$

$$= -g_l \log \Phi_G(v_l) - p_l \log \Phi_P(v_l) - e_l \log \Phi_E(v_l) \quad (3.31)$$

$$= -\tilde{g}_l \log \Phi_G(v_l) - \tilde{p}_l \log \Phi_G(-v_l) \quad (3.32)$$

nous rappelons que $\tilde{g}_l = g_l + e_l/2$, $\tilde{p}_l = p_l + e_l/2$, et le gradient est

$$\nabla_{\boldsymbol{\theta}} L_l(\boldsymbol{\theta}) = -e_l(v_l) \boldsymbol{x}_l \quad (3.33)$$

où

$$e_l(v_l) = \tilde{g}_l \psi_\kappa(v_l) + (\tilde{g}_l - 1) \psi_\kappa(-v_l) \quad (3.34)$$

$$\psi_\kappa(v) = \frac{\Phi'_\kappa(v)}{\Phi_\kappa(v)} = \frac{1}{\sigma^*} \frac{10^{-0.5v/\sigma} + \frac{1}{2}\kappa}{10^{0.5v/\sigma} + 10^{-0.5v/\sigma} + \kappa} = \frac{1}{\sigma^*} F_\kappa(-v), \quad (3.35)$$

où comme précédemment $\sigma^* = \sigma \log_{10} e$, et nous définissons

$$F_\kappa(v) = \frac{10^{0.5v/\sigma} + \frac{1}{2}\kappa}{10^{0.5v/\sigma} + 10^{-0.5v/\sigma} + \kappa} = 1 - F_\kappa(-v), \quad (3.36)$$

et donc

$$\epsilon_l(v_l) = \frac{1}{\sigma^*} \left(\tilde{g}_l F_\kappa(-v_l) + (\tilde{g}_l - 1) F_\kappa(v_l) \right) \quad (3.37)$$

$$= \frac{1}{\sigma^*} \left(\tilde{g}_l - F_\kappa(v_l) \right). \quad (3.38)$$

En utilisant (3.38) dans (2.23) nous obtenons la même équation que dans (3.15) sauf que $\Phi(v)$ doit être remplacé par $F_\kappa(v)$ et le terme multiplicatif $\frac{1}{\sigma^*}$ doit être absorbé par le pas d'adaptation K . Ceci nous donne le nouvel algorithme de notation κ -Elo

$$\hat{\theta}_{n+1,i} = \hat{\theta}_{n,i} + K[s_i - F_\kappa(\Delta_i)], \quad (3.39)$$

où comme précédemment i) $\Delta_i = \hat{\theta}_i - \hat{\theta}_j$ (j est l'index de l'équipe adverse de i), ii) comme dans l'algorithme Elo, K est le taux d'apprentissage qui règle la croissance/décroissance maximale des notes, iii) $s_i \in \left\{0, \frac{1}{2}, 1\right\}$ est une variable indicatrice du résultat du match (le score), et iv) $F_\kappa(\Delta_i)$ est la moyenne du score s_i ; en effet en utilisant (3.26)-(3.28) dans (3.36)

$$F_\kappa(\Delta_i) = \Pr \left\{ i \succ j | \hat{\theta}_i, \hat{\theta}_j \right\} + \frac{1}{2} \Pr \left\{ i \doteq j | \hat{\theta}_i, \hat{\theta}_j \right\} = \mathbb{E}_{Y|\hat{\theta}_i, \hat{\theta}_j} [s_i(Y)]. \quad (3.40)$$

Le nouvel algorithme κ -Elo garde l'interprétation et la simplicité de l'algorithme Elo tout en nous permettant de modéliser la relation entre le gain et l'égalité à travers le paramètre $\kappa \geq 0$. Nous rappelons que l'algorithme Elo n'est rien d'autre qu'une version particulière de κ -Elo pour $\kappa = 2$. Nous allons présenter des exemples numériques dans la section 4 afin d'illustrer ces propriétés.

3.2.2 κ dans κ -Elo

Le paramètre κ dans (3.39) est utilisé pour que le modèle s'ajuste au taux d'égalité entre les deux équipes adverses. Sommes-nous capables de nous prononcer sur le paramètre κ avant l'évaluation expérimentale de l'algorithme (3.39)?

Si nous supposons que le modèle représente parfaitement les données, et que le processus génératif des résultats des matchs est ergodique, nous pouvons estimer les probabilités moyennes empiriques des événements gain/perte/égalité à partir de (3.26)-(3.28) de la manière suivante:

$$\bar{p}_G = \frac{1}{N} \sum_{l=1}^N \mathbb{1}[y_l = 1], \quad \bar{p}_P = \frac{1}{N} \sum_{l=1}^N \mathbb{1}[y_l = -1], \quad \bar{p}_E = \frac{1}{N} \sum_{l=1}^N \mathbb{1}[y_l = 0], \quad (3.41)$$

et, en utilisant, (3.28) nous avons

$$\bar{p}_E \approx \kappa \sqrt{\bar{p}_G \bar{p}_P}. \quad (3.42)$$

Nous introduisons la variable intermédiaire $\bar{\delta}$ représentative de la différence des fréquences des gains et des pertes $\bar{\delta} = \bar{p}_G - \bar{p}_P$, et en appliquant la loi totale des probabilités (empirique), $\bar{p}_G = \frac{1}{2}(1 - \bar{p}_E + \bar{\delta})$ et $\bar{p}_P = \frac{1}{2}(1 - \bar{p}_E - \bar{\delta})$ nous obtenons

$$\bar{p}_E \approx \frac{\kappa}{2} \sqrt{(1 - \bar{p}_E)^2 - \bar{\delta}^2}. \quad (3.43)$$

Ainsi, pour une valeur relativement petite de la différence des fréquences $\bar{\delta}$ (par exemple $\bar{\delta} = 0.2$), le terme $\bar{\delta}^2$ est négligeable et la relation entre la probabilité empirique des égalités \bar{p}_E et κ peut se simplifier de la manière suivante:

$$\bar{p}_E \approx \frac{\kappa}{2 + \kappa}. \quad (3.44)$$

Donc, l'utilisation de κ -Elo avec $\kappa = 2$ (qui correspond à l'implémentation de l'algorithme Elo), implique que $\bar{p}_E = 0.5$. Comme aucun des sports de compétition dans lesquels l'algorithme Elo est

appliqué ne valide cette supposition, nous nous attendons à ce que le nouvel algorithme κ -Elo (avec une valeur plus appropriée de κ) améliore l'estimation des probabilités des gains, pertes et égalités.

En utilisant (3.44), nous pouvons aussi estimer la valeur adéquate de κ de la manière suivante

$$\bar{\kappa} \approx \frac{2\bar{p}_E}{1 - \bar{p}_E}. \quad (3.45)$$

Par exemple en utilisant $\bar{p}_E = 0.26$ (qui est la fréquence des égalités de la saison 2017 – 2018 du championnat de football anglais, voir la figure 4.4) nous obtenons $\bar{\kappa} \approx 0.7$.

Est-ce que cette valeur est acceptable?

Avant de répondre à cette question, nous devons signaler un problème conceptuel qui peut survenir dans la modélisation des égalités. En fait, dans la littérature, les deux tentatives d'inclusion de l'égalité (la méthode de seuillage (3.25) et la méthode que nous utilisons (3.26)-(3.28)) n'implémentent aucune contrainte sur la relation entre les valeurs prédites de probabilités à part la symétrie $\Phi_G(v) = \Phi_P(v)$ et la loi totale de probabilité $\Phi_G(v) + \Phi_P(v) + \Phi_E(v) = 1$. Ce qui fait que des scénarios contre-intuitifs peuvent avoir lieu: supposons que l'équipe i confronte l'équipe j avec $\varepsilon = \hat{\theta}_i - \hat{\theta}_j$ (avec $\varepsilon > 0$ une petite différence des notes), nous avons $\Phi_G(\varepsilon) \geq \Phi_P(\varepsilon)$ et notre intuition est comme suit: le gain de l'équipe la plus forte est plus probable que sa perte. Par contre, nous ne pouvons pas nous prononcer sur la probabilité d'égalité. Devons-nous nous attendre à ce que la probabilité de l'égalité soit plus grande que la probabilité de gain ou de perte? Rien ne contraint notre modèle (ni le modèle (3.25)) de ne pas produire $\Phi_G(\varepsilon) = 0.41$, $\Phi_P(\varepsilon) = 0.36$ et $\Phi_E(\varepsilon) = 0.23$ et l'interprétation serait: la perte de l'équipe la plus forte est plus probable que l'égalité. Ce qui est totalement contre-intuitif.

Il serait donc plus logique d'enlever ce type de résultat de l'espace des solutions: pour les équipes à notes égales, nous imposons que la probabilité d'égalité soit plus grande que la probabilité de gain ou de perte, ce qui nous oblige à utiliser une valeur de κ qui satisfait la condition suivante

$$\Phi_E(\varepsilon) \geq \Phi_G(\varepsilon) \iff \kappa \geq 1 \quad (3.46)$$

où cette dernière inégalité découle du modèle (3.26)-(3.28). Ceci constitue une restriction assez importante, qui sous-entend que nous sommes forcés de modéliser les égalités avec une (grande) fréquence ($\bar{p}_E > 0.33$, voir (3.44)).

Bien que l'utilisation d'un tel modèle inadapté n'est pas optimale, nous ne savons pas à quel point cela influence la capacité de prédiction. Nous rappelons que la version courante de l'algorithme Elo utilise $\kappa = 2$, qui correspond à la modélisation des égalités avec une fréquence $\bar{p}_E = 0.5$. Nous allons explorer l'impact de cette restriction en évaluant la capacité de prédiction correspondante sur des exemples numériques.

Chapitre 4

Exemple numérique

Afin d'illustrer la capacité de prédiction de l'algorithme κ -Elo pour différentes valeurs de κ , nous allons commencer par l'appliquer sur les résultats du championnat de football anglais disponible sur (Football-data.co.uk (2020)), ensuite nous allons l'appliquer sur la ligue nationale de hockey (sportsbookreviewsonline.com (2020)). Le titre du championnat de la première ligue anglaise est disputé par $M = 20$ équipes; chacune d'entre elles joue deux matchs contre toutes les équipes restantes: un à domicile et l'autre sur le territoire de l'équipe adverse. Nous allons prendre en considération une saison à la fois, ce qui fait que $n = 1, \dots, N$ est l'indice des matchs dans l'ordre chronologique, avec $N = M(M - 1) = 380$ le nombre total de rencontres.

Le football et le hockey sont des sports reconnus pour donner l'avantage à l'équipe jouant « à domicile »; une équipe a plus de chance de gagner à domicile que chez l'adversaire, ceci peut être expliqué par la croissance de la motivation des joueurs sous l'influence des supporters. Dans le cadre de la notation, ceci peut être modélisé en introduisant un paramètre $\eta \geq 0$ qui fait croître d'une manière artificielle la note de l'équipe qui joue à domicile. Cette croissance correspond à la translation de la probabilité conditionnelle de la manière suivante:

$$\Phi_G^{aad}(v) = \Phi_G(v + \eta\sigma), \quad \Phi_P^{aad}(v) = \Phi_P(v + \eta\sigma), \quad \Phi_E^{aad}(v) = \Phi_E(v + \eta\sigma), \quad (4.1)$$

nous rappelons que $v = \theta_i - \theta_j$, la différence des notes des deux équipes adverses, et nous supposons que i est l'équipe qui joue à domicile. La valeur du paramètre de l'avantage à domicile η doit être

choisie convenablement afin d'améliorer la capacité de prédiction de l'algorithme. Nous présentons dans la figure 4.1 les probabilités de prédiction pour différentes valeurs de η .

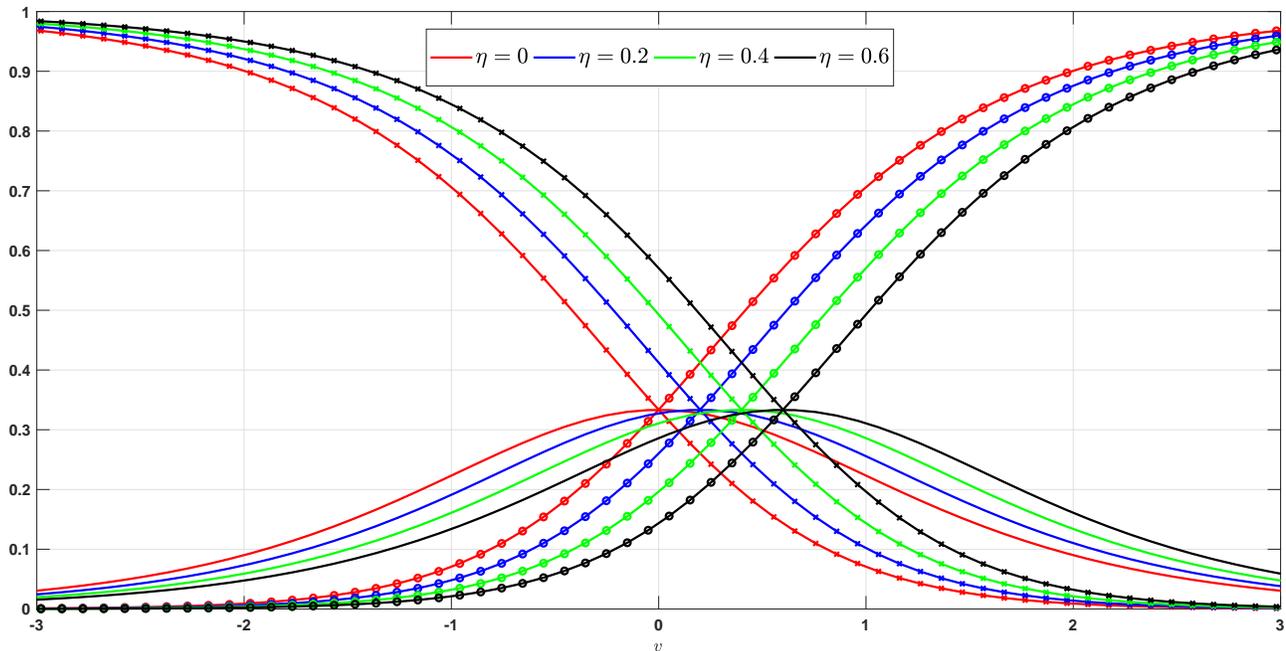


Figure 4.1 – La forme des fonctions de prédiction de κ -Elo pour différentes valeurs de η , avec $\kappa = 1$ et $\sigma = 1$, la probabilité de gain $\Phi_G^{aad}(v)$ est marquée par 'o'; la fonction de perte $\Phi_P^{aad}(v)$ est marquée par 'x' et la fonction d'égalité $\Phi_E^{aad}(v)$ est la ligne sans marque.

Tout comme pour l'algorithme de notation de la FIFA, FIFA (2019), nous fixons $\sigma = 600$. Nous initialisons les notes à zéro $\theta_{0,m} = 0$. Comme expliqué précédemment, ces valeurs sont arbitraires et n'influent pas sur les prédictions finales. Dans tout ce qui suit, afin d'éliminer la dépendance du facteur d'échelle σ , nous allons utiliser la normalisation $K = \sigma \tilde{K}$: pour une valeur donnée de \tilde{K} la prédiction sera la même indépendamment du facteur d'échelle σ .

Dans la figure 4.2, nous présentons un exemple des valeurs des notes estimées $\hat{\theta}_{n,m}$ de quelques équipes. Nous remarquons qu'une grande partie des matchs (à partir du début de la saison) sont dédiés à la convergence de l'algorithme, c'est la période «d'apprentissage» de l'algorithme. Bien évidemment nous pouvons accélérer l'apprentissage en augmentant la valeur du pas \tilde{K} , mais ceci augmentera la variabilité des notes estimées. Ce compromis est étroitement relié à l'application du gradient stochastique et son étude ne fait pas partie de nos objectifs; nous le mentionnons pour argumenter le fait que nous allons juste utiliser la deuxième moitié de la saison pour l'évaluation de la capacité de prédiction. Nous supposons que l'algorithme finit de converger dans la première

moitié de la saison et que les notes suivent les performances des équipes. Cette division est en effet arbitraire, mais notre objectif ici est d'explorer l'effet du paramètre d'égalité κ sur la capacité de prédiction et non pas la résolution du compromis du gradient stochastique.

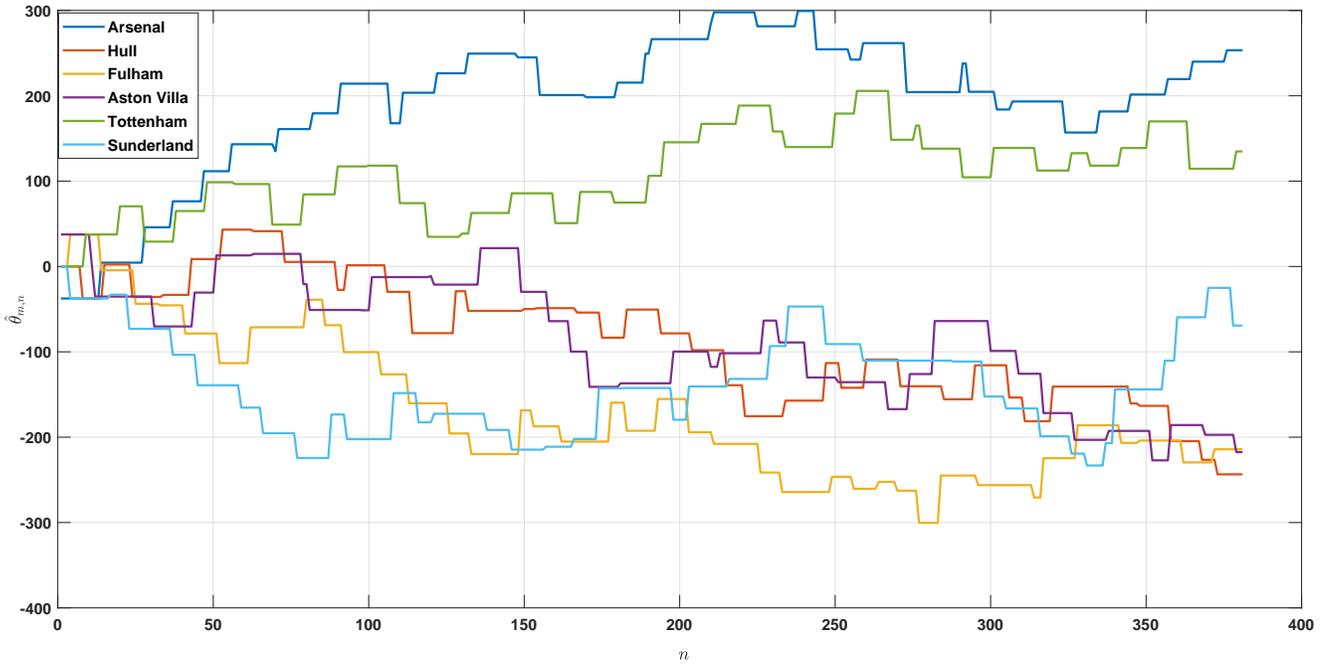


Figure 4.2 – Évolution des notes $\hat{\theta}_{m,n}$ de quelques équipes de la première ligue anglaise de football au cours de la saison 2013 – 2014; $N = 380$, $\sigma = 600$, $\bar{K} = 0.125$, $\eta = 0.3$, $\kappa = 0.7$. Nous supposons que, la première moitié de la saison (les 190 premier matchs) est suffisante pour absorber la phase d'apprentissage, et la seconde moitié est libre de l'effet de l'initialisation.

Nous rappelons que la prédiction du résultat d'un match futur $\{G, P, E\}$ (respectivement gain, perte et égalité), n'est rien d'autre qu'une estimation de la probabilité du résultat du match qui se déroule à l'instant l en utilisant les notes $\hat{\theta}_{l-1}$ disponibles à l'instant $l - 1$:

$$\hat{p}_{l,G} = \Phi_G(\mathbf{x}_l^T \hat{\theta}_{l-1}), \quad \hat{p}_{l,P} = \Phi_P(\mathbf{x}_l^T \hat{\theta}_{l-1}), \quad \hat{p}_{l,E} = \Phi_E(\mathbf{x}_l^T \hat{\theta}_{l-1}). \quad (4.2)$$

Le score logarithmique est un score de prédiction propre Gneiting & Raftery (2007): sa moyenne n'est maximisée que pour la vraie distribution de probabilité. Pour l'évaluation de la capacité de prédiction, nous avons utilisé le négative de la moyenne du score logarithmique Gelman & Vehtari (2014) sur la deuxième moitié de la saison de la façon suivante:

$$\overline{LS} = \frac{2}{N} \sum_{l=N/2+1}^N LS_l \quad (4.3)$$

avec

$$LS_l = -(g_l \log \hat{p}_{l,G} + p_l \log \hat{p}_{l,P} + e_l \log \hat{p}_{l,E}) \quad (4.4)$$

Comme expliqué auparavant, la prédiction d'un match futur en utilisant l'algorithme Elo conventionnel n'est pas formellement spécifiée. Dans le cadre de l'algorithme Elo binaire ($\kappa = 0$), la probabilité d'égalité est nulle $\hat{p}_{l,E} = 0$, ce qui fait que le score logarithmique divergera à l'infini. Comme variante de l'algorithme Elo conventionnel, nous allons plutôt utiliser l'approche heuristique de Lasek *et al.* (2013) qui consiste à utiliser l'algorithme Elo ternaire conventionnel $\kappa = 2$ pour l'inférence des notes tout en basant la prédiction sur le modèle utilisé par l'algorithme κ -Elo avec une valeur différente de $\kappa = \check{\kappa}$. Ceci est évidemment une non-correspondance entre l'estimation et la prédiction. La méthode de Lasek *et al.* (2013) propose d'utiliser $\check{\kappa} = 1$ (la valeur minimale qui garantit (3.46)), ce qui revient à supposer une fréquence d'égalité $\bar{p}_E = 0.33$.

Nous montrons dans les figures 4.3 et 4.4 le score logarithmique \overline{LS} (respectivement de la saison 2013 – 2014 et la saison 2017 – 2018) pour différentes valeurs du paramètre d'égalités κ , et pour différentes valeurs du paramètre de l'avantage à domicile η . Nous comparons nos prédictions avec celles basées sur les cotes des paris du site Bets365 (disponible sur Football-data.co.uk (2020)) que nous calculons de la façon suivante:

$$\tilde{p}_G \propto 1/o_G, \quad \tilde{p}_P \propto 1/o_P, \quad \tilde{p}_E \propto 1/o_E \quad (4.5)$$

où \propto est le symbole de proportionnalité, o_G , o_P et o_E sont respectivement les cotes des paris de gain, perte et de l'égalité de l'équipe à domicile. Ces valeurs de probabilité sont ensuite normalisées pour qu'elles s'additionnent à un. La valeur du score logarithmique associé à ces prédictions constitue une ligne de référence par rapport à laquelle nous comparons la performance de l'algorithme κ -Elo.

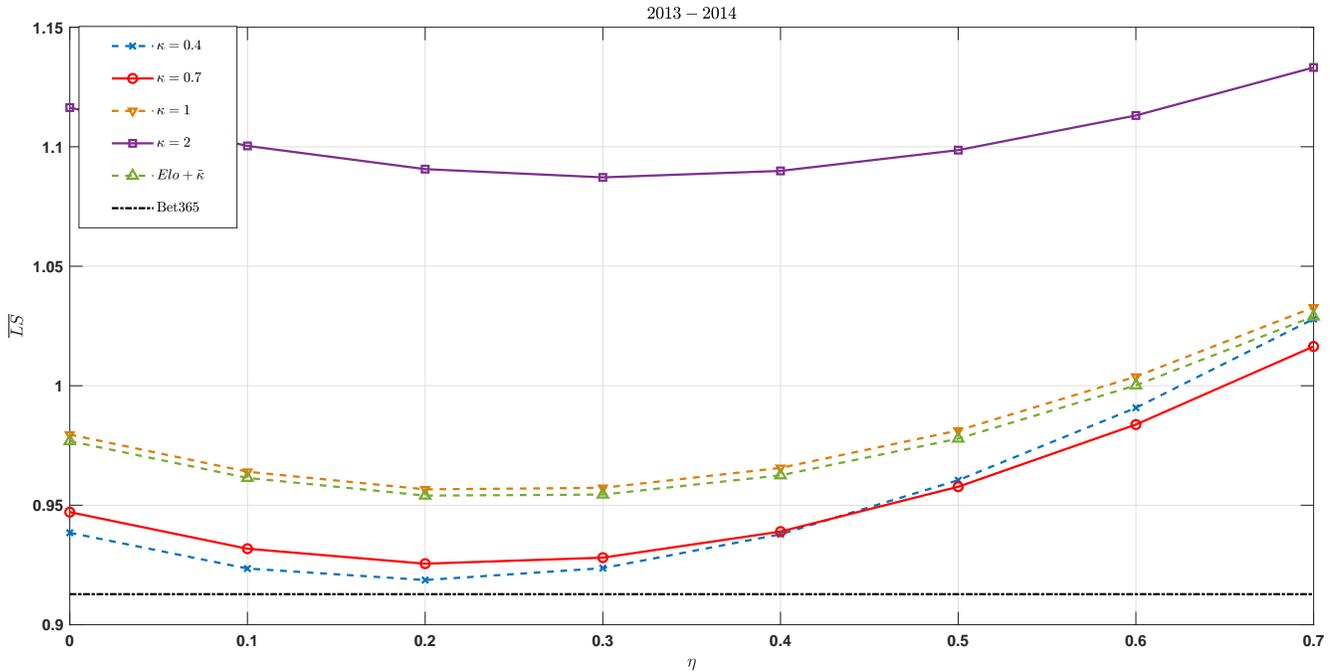


Figure 4.3 – Score logarithmique 4.3 de la saison 2013 – 2014 de la première ligue anglaise de football en fonction du paramètre de l’avantage à domicile η , le score est calculé sur la deuxième moitié de la saison avec $\sigma = 600$, $\tilde{K} = 0.125$, les différentes valeurs de κ apparaissent dans la légende. La fréquence d’égalité est $p_E = 0.17$ qui correspond à $\bar{\kappa} = 0.4$. Le résultat "Elo + κ " est obtenu en utilisant les notes de l’algorithme Elo conventionnel tout en utilisant $\tilde{\kappa} = 1$ pour la prédiction. Le résultat "Bet365" est basé sur les cotes des paris.

Nous observons que l’introduction du paramètre κ améliore le score logarithmique de notre prédiction des matchs de football. Par contre, l’utilisation de l’algorithme Elo ternaire ($\kappa = 2$) avec sa prédiction correspondante donne un score logarithmique pire que celui obtenu en utilisant une prédiction non adaptée ($\tilde{\kappa} = 1$ selon la méthode de Lasek *et al.* (2013)). Les résultats obtenus par la méthode de Lasek *et al.* (2013) sont très proches de ceux obtenus en utilisant l’algorithme κ -Elo avec $\kappa = 1$. Nous remarquons que pour la saison 2013 – 2014, dans laquelle la fréquence d’égalité est faible, l’utilisation d’une valeur de $\kappa = \bar{\kappa}$ produit un score logarithmique bien meilleur que celui obtenu en utilisant une valeur plus grande ($\kappa = 1$ ou $\kappa = 2$).

Afin d’explorer davantage l’effet du paramètre κ , nous allons présenter les résultats numériques obtenus pour la saison 2017 – 2018 de la saison régulière de la ligue nationale de hockey sportsbookreviewsonline.com (2020). La structure de la ligue nationale de hockey diffère de la première ligue anglaise. Le championnat est disputé par $M = 31$ équipes qui s’affrontent dans une saison régulière suivie d’une série éliminatoire. Les équipes sont réparties sur deux associations, et chaque

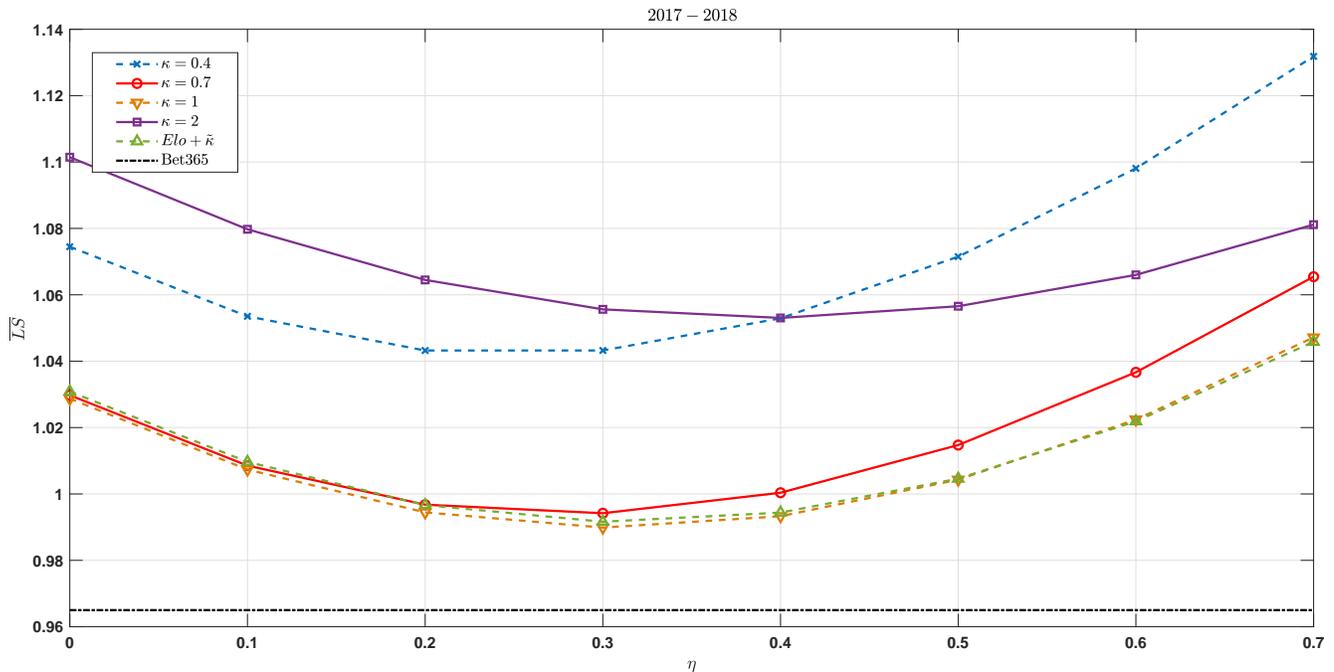


Figure 4.4 – Score logarithmique 4.3 de la saison 2017 – 2018 de la première ligue anglaise de football en fonction du paramètre de l’avantage à domicile η , le score est calculé sur la deuxième moitié de la saison avec $\sigma = 600$, $\tilde{K} = 0.125$, les différentes valeurs de κ apparaissent dans la légende. La fréquence d’égalité est $\bar{p}_E = 0.26$ qui correspond à $\bar{\kappa} = 0.7$. Le résultat "Elo + κ " est obtenu en utilisant les notes de l’algorithme Elo conventionnel tout en utilisant $\tilde{\kappa} = 1$ pour la prédiction. Le résultat "Bet365" est basé sur les cotes des paris.

association se compose de deux divisions. Chaque équipe joue 41 matchs à domicile et 41 matchs à l’extérieur; 4 ou 5 matchs contre les autres équipes de la même division, 3 matchs face aux équipes de l’autre division mais de la même association, et 2 rencontres contre les équipes de l’autre association. Comme il n’y a pas d’égalité dans la ligue nationale de hockey, nous considérons le passage aux prolongations ou aux tirs de pénalité comme des égalités. Comme précédemment, nous utilisons $n = 1, \dots, N$ comme indice chronologique des rencontres avec $N = \prod 41M = 1271$, et nous utilisons juste la moitié de la saison régulière pour l’évaluation (tout comme le football, la première moitié est dédiée à l’apprentissage), voir figure 4.5.

Nous montrons sur la figure 4.6 le score logarithmique \overline{LS} de la saison 2017 pour différentes valeurs du paramètre d’égalités κ , et pour différentes valeurs du paramètre de l’avantage à domicile η . À notre connaissance, aucun bookmaker n’offre des paris pour le passage aux prolongations ou aux pénalités, ce qui fait que nous n’avons pas de ligne de référence comme précédemment. Nous remarquons que l’utilisation de $\kappa = \bar{\kappa}$ donne la meilleure performance, et l’algorithme Elo ($\kappa = 2$)

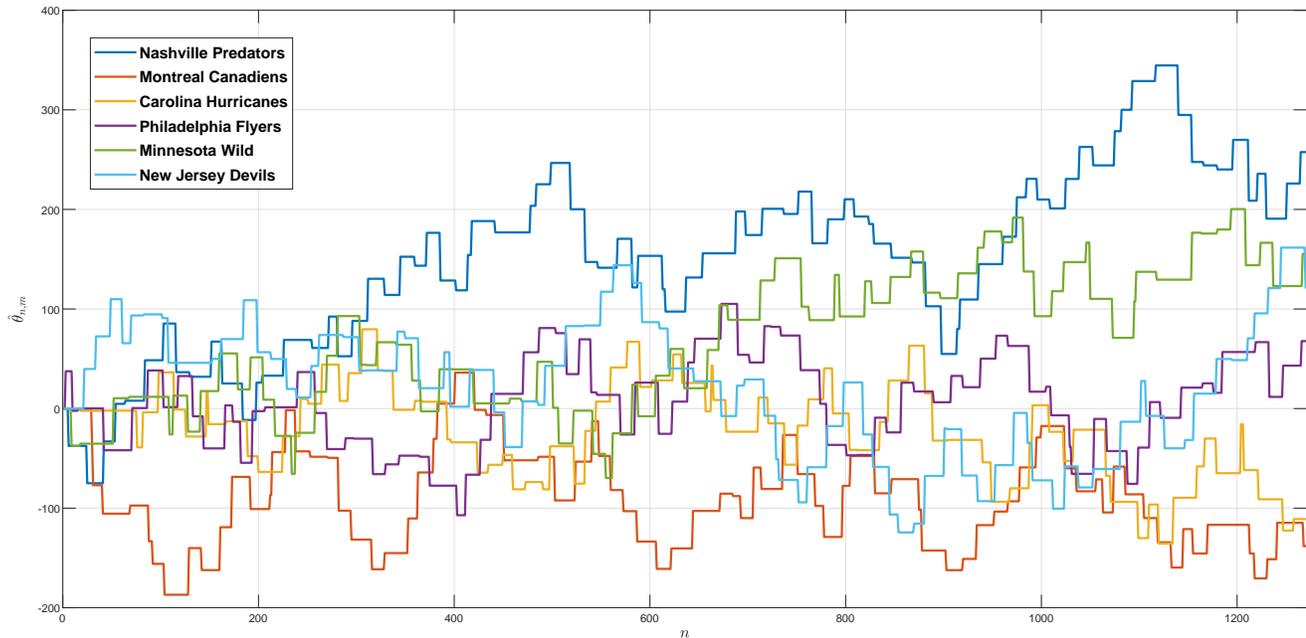


Figure 4.5 – Évolution des notes $\hat{\theta}_{m,n}$ de quelques équipes de la ligue nationale de hockey au cours de la saison 2017; $N = 1271$, $\sigma = 600$, $\tilde{K} = 0.125$, $\eta = 0.3$, $\kappa = 0.7$. Nous supposons que la première moitié de la saison suffit pour absorber la phase d'apprentissage, et la seconde moitié est libre de l'effet de l'initialisation.

donne la pire performance. Tout comme précédemment, l'utilisation de l'algorithme Elo ($\kappa = 2$) avec sa prédiction correspondante donne une performance pire que celle obtenue en utilisant une prédiction non adaptée ($\tilde{\kappa} = 1$) selon la méthode de Lasek *et al.* (2013), et la différence entre cette dernière et κ -Elo avec $\kappa = 1$ est minime.

La synthèse est comme suit: les résultats obtenus sur ces données expérimentales de football et de hockey sont similaires, et ils confirment notre discussion théorique précédente. Comme nous nous y attendions, l'introduction du paramètre d'égalité κ ne change pas la performance de l'algorithme Elo d'une façon dramatique. Cependant, une amélioration est obtenue en utilisant $\kappa = 1$. Cette recommandation est motivée par la discussion dans la section 3.2.2 et n'introduit aucune complexité d'implémentation supplémentaire.

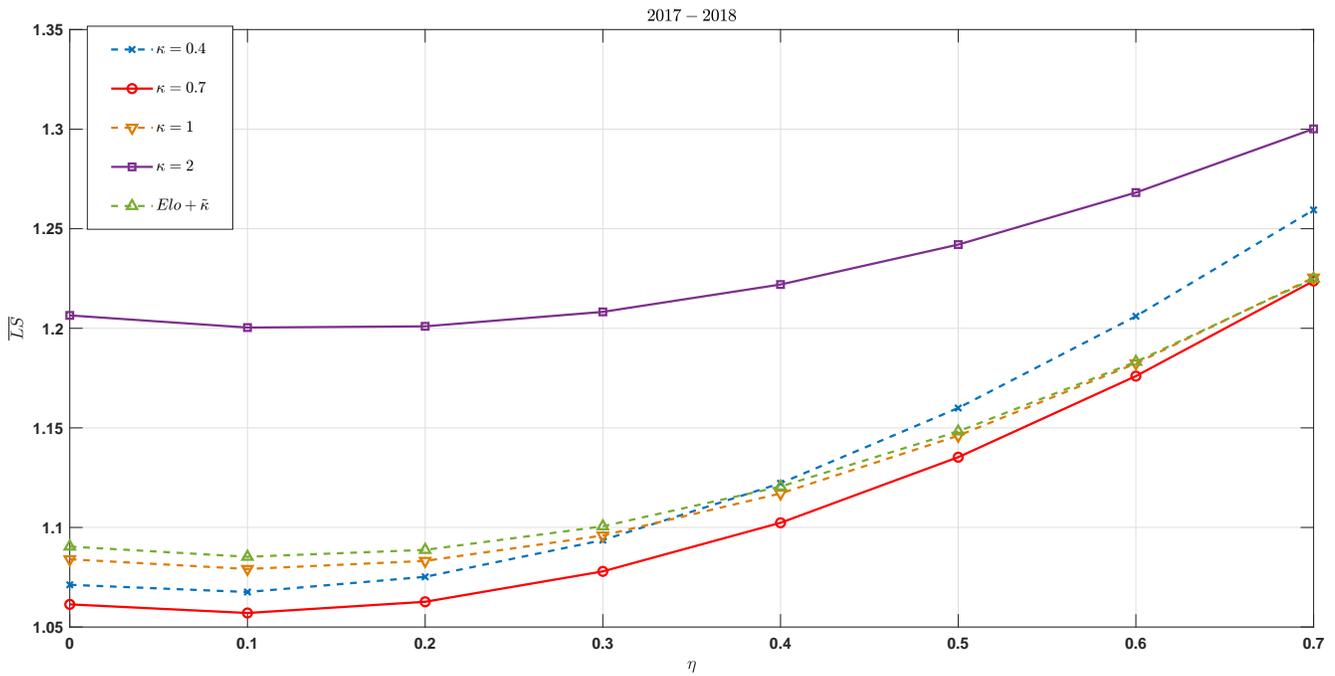


Figure 4.6 – Score logarithmique 4.3 de la saison 2017 – 2018 de la première ligue nationale de hockey en fonction du paramètre de l'avantage à domicile η , le score est calculé sur la deuxième moitié de la saison avec $\sigma = 600$, $\tilde{K} = 0.125$, les différentes valeurs de κ apparaissent dans la légende. La fréquence d'égalité est $p_E = 0.23$ qui correspond à $\bar{\kappa} = 0.59$. le résultat "Elo + $\bar{\kappa}$ " est obtenu en utilisant les notes de l'algorithme Elo conventionnel tout en utilisant $\check{\kappa} = 1$ pour la prédiction.

Chapitre 5

Conclusion et extensions possibles

Ce mémoire a été principalement consacré à l'explication de la logique et la fondation mathématique derrière l'algorithme Elo. Nos travaux peuvent être résumés en ce qui suit:

- Nous avons expliqué le fait que l'algorithme Elo, dans sa version binaire, n'est autre que le fruit de l'application du gradient descendant stochastique pour la résolution du problème d'estimation du maximum de vraisemblance. Ceci a déjà été trouvé par Király & Qian (2017), et nous l'avons présenté en détail comme une base à partir de laquelle nous élaborons le reste du rapport.
- Nous avons montré que dans sa version originale gaussienne, l'algorithme Elo résout d'une manière approximative le problème de maximisation de la vraisemblance.
- Nous avons montré la modélisation implicite de l'algorithme Elo avec égalité. Bien que l'algorithme est utilisé depuis des décennies, la modélisation de l'égalité n'est pas présentée dans la littérature, ce qui fait qu'il n'y a pas une formulation explicite de la prédiction des égalités. Nous avons comblé cette lacune.
- Nous avons utilisé le modèle de Davidson (1970) pour la généralisation de l'algorithme Elo; nous appelons le nouvel algorithme κ -Elo, il garde la même simplicité de l'algorithme Elo tout en nous permettant d'ajuster la fréquence des égalités à travers un paramètre additionnel κ . De cette façon, nous avons démontré le fait que l'algorithme Elo n'est qu'un cas particulier de κ -Elo et qu'il suppose d'une manière implicite que la fréquence des égalités est de 50%.

- Nous avons abordé la contrainte dans la relation entre la valeur de la probabilité de l'égalité et de la perte des équipes à notes similaire ; nous postulons que dans un cas pareil, la probabilité d'égalité doit être supérieure à la probabilité de gain/perte. Bien que cette contrainte est absente dans la littérature, nous pensons qu'elle mérite plus d'attention afin de construire des modèles et des algorithmes plus adaptés à la notation. Dans notre cas, ces contraintes nous imposent l'utilisation de l'algorithme κ -Elo avec une valeur de $\kappa \geq 1$, ce qui revient à supposer une fréquence d'égalité $\bar{p}_E \geq 0.33$. Ceci est clairement une limitation dans le cas où les fréquences d'égalités des jeux en question sont inférieures à 33%.
- Afin d'illustrer le concept général, nous avons présenté des exemples numériques sur le championnat de football anglais et sur le championnat nationale de hockey.
- Finalement, nous concluons que, bien que l'algorithme Elo ait satisfait le besoin d'un système de notation simple depuis plusieurs décennies, il est encore possible de concevoir des systèmes de notation qui sont plus performants et plus flexibles, tout en étant aussi simples. En particulier, l'algorithme κ -Elo que nous proposons est meilleur dans le sens qu'il prend en considération les fréquences des égalités tout en ayant une complexité très faible.

Le futur suivi de ce travail pourra inclure ce qui suit:

- L'inclusion d'autres paramètres significatifs (tel que le score de la rencontre).
- L'étude de l'impact de la valeur du pas d'apprentissage sur la capacité de prédiction dans différents sports de compétition.
- L'implémentation des contraintes moins restrictives (que $\kappa > 1$) sur la relation entre les probabilités prédites en cas de notes similaires.
- L'utilisation d'un paramètre d'avantage à domicile qui soit spécifique pour chaque équipe, et envisager la possibilité de l'inférer à partir des données.

Références

- Bagnoli M & Bergstrom T (2005). Log-concave probability and its applications. *Economic Theory* 26: 445.
- Bradley RA & Terry ME (1952). Rank analysis of incomplete block designs: 1 the method of paired comparisons]. *Biometrika* 39 324–345.
- Caron F & Doucet A (2012). Efficient Bayesian Inference for Generalized Bradley-Terry Models . *Journal of Computational and Graphical Statistics*, 21, 174-196.
- David HA (1988). The Method of Paired Comparisons. 2nd ed. *Griffin's Statistical Monographs and Courses* 41. Griffin, London. MR0947340.
- Davidson RR (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65, 317–328, URL <http://www.jstor.org/stable/2283595>.
- Elo AE (2008). The Rating of Chess Players, Past and Present. *Ishi Press International*.
- Fahrmeir L & Tutz G (2006). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association*, 89, 1438–1449, URL <http://dx.doi.org/10.1093/biomet/39.3-4.324>.
- Farhang-Boroujeny B (1998). Adaptive filters, theory and applications. *John Wiley and Sons*.
- FIFA (2019). Fédération Internationale de Football Association: men's ranking procedure. URL <https://www.fifa.com/fifa-world-ranking/procedure/men>.
- Football-data.co.uk (2020). Historical football results and betting odds data. URL <https://www.football-data.co.uk/data.php>.
- Gelman, A. JH & Vehtari A (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24, 997–1016, URL <https://doi.org/10.1007/s11222-013-9416-2>.
- Glickman M & Hennessy J (2015). A stochastic rank ordered logit model for rating multi-competitor games and sports. *Journal of Quantitative Analysis in Sports*, 11. DOI:10.1515/jqas-2015-0012.
- Glickman ME (1999). Parameter estimation in large dynamic paired comparison experiments. *Applied Statistics*, 48:377–394.
- Gneiting T & Raftery AE (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of American Statistical Association* 102, 359-378.

- Herbrich R & Graepel T (2006). Trueskill(tm): A bayesian skill rating system. *Technical report*, URL <https://www.microsoft.com/en-us/research/publication/trueskilltm-a-bayesian-skill-rating-system-2/>.
- Király FJ & Qian Z (2017). Modelling Competitive Sports: Bradley- Terry-Elo Models for Supervised and On-Line Learning of Paired Competition Outcomes. *arXiv e-prints*, *arXiv:1701.08055*.
- Langville AN & Meyer CD (2012). *Who's 1, The Science of Rating and Ranking*. Princeton University Press.
- Lasek J, Szlávik Z & Bhulai S (2013). The predictive power of ranking systems in association football. *IJAPR*, 1:27–46.
- Rao PV & Kupper LL (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model,. *Journal of the American Statistical Association*, 62, 194–204.
- Rumelhart DE, Hinton GE & Williams RJ (2012). Learning representations by back-propagating errors . *Nature*, 323, 533-536, URL <https://doi.org/10.1038/323533a0>.
- sportsbookreviewsonline.com (2020). Historical hockey results . URL <https://www.sportsbookreviewsonline.com/scoresoddsarchives/nhl/nhloddsarchives.htm>.
- Thurstone LL (1927). A law of comparative judgement. *Psychological Review*, 34, 273-286.
- Toulis P & Airoldi EM (2014). Asymptotic and finite-sample properties of estimators based on stochastic gradients. *arXiv e-prints*1408.2923.
- Wikipedia (2019). Wikipedia: Elo rating system . URL https://en.wikipedia.org/wiki/Elo_rating_system.
- Zermelo E (1929). "Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung [The calculation of tournament results as maximum likelihood problem]". *Mathematische Zeitschrift*, 29, 436-460.