SOCIETE FRANÇAISE DE STATISTIQUE

Journal de la Société Française de Statistique
Vol. 156 No. 2 (2015)

Autour des procédures de classement : exemples du tennis, du tennis de table et du golf

Title: Around the ranking process : examples of tennis, table tennis and golf

Loïc Champagne[1] et Léo Gerville-Réache[1]

Résumé : En sport individuel, les sportifs cherchent à évaluer le niveau de leur pratique et parviennent à obtenir un classement en participant à des compétitions officielles. Le classement a pour but d'estimer le niveau de pratique des sportifs dans leur sport. Ce principe s'applique à tous les sports et ainsi, chaque joueur peut voir évoluer son niveau et le comparer à celui de ses partenaires ou adversaires. Pour autant, chaque procédure de classement se doit de s'adapter au sport auquel elle s'applique. Nous essayons ici de comprendre la logique de ces procédures et tentons d'apporter notre contribution à l'amélioration du rapport entre le classement et le niveau de jeu. Le but de cet article est de sensibiliser, via quelques exemples, à l'utilité des approches statistiques et probabilistes dans les réflexions liées au concept de classement en sport individuel.

Abstract: In individual sport, players in sport try to estimate their level in their sport practice and succeed in obtaining a classification by participating in official competitions. The aim of ranking is to define the level of practice of the sportsmen in every sport. This postulate applies to all the sports and so, every player can see evolving his level and compare it to his partners or opponents level. However, each ranking process must adapt to the sport which it applies. Here, we try to understand the logic of these procedures, and try to bring our contribution to the improving of the relationship between ranking and game level. The purpose of this communication is to educate, with few examples, usefulness of statistical and probabilistic approaches in the reflections related to the concept of ranking in individual sport.

Mots-clés : classement, tennis, tennis de table, golf, niveau de jeu
Keywords: ranking, tennis, table tennis, golf, game level
Classification AMS 2000 : 46N30, 62F07

1. Introduction

Dans les pratiques sportives telles que le tennis, le tennis de table ou encore le golf, le niveau d'un joueur est une notion essentielle, notamment en compétition. Le classement va permettre aux joueurs d'évaluer leur niveau et celui des autres. Aussi, les compétitions sont généralement organisées par catégorie de niveau. Au tennis de table, lors d'un tournoi, plusieurs "tableaux", avec une limite de classement pour chacun, sont proposés aux joueurs (un joueur classé 1950 points ne pourra pas participer à un tableau <1500 points). Au tennis, chaque joueur se situe à l'intérieur d'une "classe de niveau" appelée série : 4ème série (classements de NC à 30/1), 3ème série (classements de 30 à 15/1), 2nde série (classements de 15 à -30) et 1ère série (classement du numéro 30 au numéro 1). La Fédération Française de Tennis organise alors des compétitions par équipes basées sur cette classification. Enfin, au golf, les compétitions sont divisées en séries qui

[1] Université de Bordeaux
E-mail : loic.champagne@u-bordeaux.fr et E-mail : leo.gerville-reache@u-bordeaux.fr

correspondent à des intervalles de handicap.

La mesure du niveau de jeu est une tâche délicate qui trouve une solution par le biais de l'attribution d'un classement (Coulom, 2010). En France, le choix de la procédure de classement est une prérogative de chaque fédération sportive (Brossard, 1994). La diversité des procédures et la complexité de certaines nous à conduit à engager une large réflexion sur le principe même de mesure d'un "niveau sportif". En 2008, Nicolas Paris étudie l'évolution de la procédure de classement au tennis depuis 1973. Il montre, dans sa thèse de l'université Bordeaux 2, comment se construit le classement des compétiteurs, en fonction du nombre de points dont le joueur est crédité au début de la saison (capital de départ), du nombre de matches (victoires et défaites) disputés dans la saison et du seuil de points à atteindre afin de monter (améliorer son classement), de se maintenir ou le cas échéant, de descendre (perdre un classement). Son étude démontre une extrême variabilité des exigences de niveau au cours des 40 dernières années, selon le classement auquel se situent les joueurs, afin de monter ou de se maintenir au même classement la saison suivante (Paris, 2008).

La plupart des sports ont vu leur procédure de classement évoluer en fonction de volontés plus souvent politiques que scientifiques. En effet, au tennis comme au tennis de table, des classements dits "intermédiaires" ont été créés afin d'encourager les nouveaux compétiteurs à acquérir rapidement un classement et à s'intéresser à la compétition (création récente du classement 40 au tennis et des classements 55 à 90 au tennis de table depuis les années 90). La nécéssité qu'ont les fédérations de rendre leur sport "populaire" a souvent produit des politiques de sur-classement pour les petits niveaux et de sous-classement pour les haut-niveaux. Ainsi, un joueur qui débute la compétition est presque sûr d'avoir un classement la saison suivante. Le nombre de points nécéssaires à l'obtention du premier classement est faible et permet donc de motiver les joueurs débutants.

Une approche scientifique doit permettre de formaliser les propriétés nécessaires et suffisantes d'une procédure de classement. Dans ce cadre, la formalisation probabiliste, basée sur la notion de probabilité de victoire, est le postulat retenu pour cette réflexion. Le modèle de Bradley and Terry (1952) est un des modèles probabilistes les plus connus. Sachant le classement de deux joueurs, il met en relation la probabilité de victoire avec l'écart de classement. Un tel modèle propose principalement une bijection entre "écart de classement" et "probablité de victoire". C'est à dire que l'écart de classement de deux joueurs en compétition influe sur la probabilité de victoire du match et inversement. Plus généralement, la littérature scientifique sur le classement et relativement abondante. L'un des concepts clés est celui de "comparaison par paires". Astié (1970), Courcoux and Séménou (1997), Glickman and Jensen (2003), Hallinan (2005) pour n'en citer que quelques-uns. Néanmoins, un problème persiste, celui de l'incomplétude du plan d'expérience. En effet, dans un sport comme le tennis de table qui comprend environ 200 000 licenciés, pour que chaque joueur rencontre chaque autre sur une année il faudrait que chaque joueur fasse 200 000 parties par an ! En moyenne, un joueur de tennis de table ne réalise qu'une centaine de parties à l'année. Aussi, chaque joueur ne rencontre que 0.05% des licenciés. C'est avec cette réalité d'une matrice de rencontre très creuse qu'il faut travailler. Plusieurs variables telles que le type de compétition ou la surface de jeu peuvent être utilisées lorsque l'information est connue et améliorer la précision des données observées. Cela dépend de ce que l'on veut mesurer et des informations que l'on a.

Nous supposons donc l'existence d'une relation générale entre handicap de jeu, probabilité de victoire et procédure de classement. Ce tryptique 1 évoluant autour de la notion de niveau de jeu.

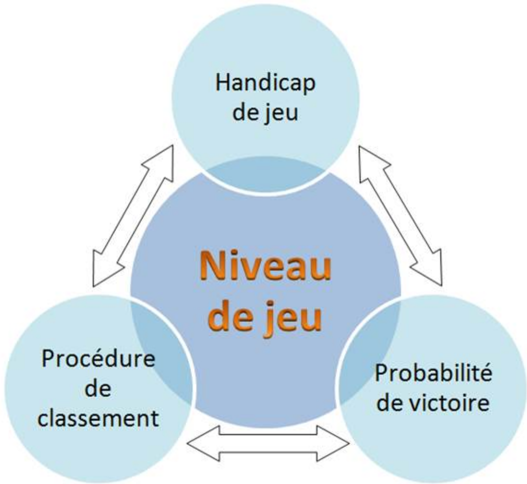

FIGURE 1. *Tryptique autour du niveau de jeu*

Les calculs nécessaires à la construction du classement sont dépendants de variables telles que le capital de points au départ de la saison, le seuil de montée et de descente, le nombre de matches joués, le nombre de victoires ou de défaites prises en compte, etc... Ces variables sont choisies, pour chaque procédure de classement, par les Fédérations sportives. Aussi, chaque sport possède sa propre procédure de classement. La notion de "probabilité de victoire" est fortement liée à celle de "handicap de jeu". La mise au point d'une règle de handicap est une étape indispensable dans tous les sports. En effet, permettre à deux joueurs, de classements différents, de se mesurer en rééquilibrant les chances de victoire de l'un et de l'autre est une nécessité pour l'entrainement comme pour certaines compétitions. Ces chances de victoire se traduisent par une probabilité de victoire et/ou de défaite pour les deux joueurs qui se rencontrent. Dès qu'une procédure de classement existe, il est possible d'en déduire une distance probabiliste sous-jacente entre tout couple de classés et de compenser cet écart de probabilité de victoire par une règle de handicap de jeu.

Le "niveau de jeu" d'un sportif (au tennis, tennis de table et golf) est une notion très subjective. Il se traduit par le classement qui correspond aux résultats de ce sportif au cours d'une saison sportive qui dure généralement 12 mois. Cependant un compétiteur n'a pas, à chaque match, le même niveau notamment en sport de raquettes. De nombreuses grandeurs incontrôlables entrent en jeu (sommeil, emploi du temps, surfaces de jeu, intéraction avec les adversaires ou encore la marque des balles) et sont sources d'incertitudes. Aussi, son niveau de jeu au moment t du match ne vérifie pas toujours le classement qui lui est attribué. On appellera alors ce classement, le classement "moyen" d'un joueur. Peut-on considérer que le classement correspond au niveau de jeu d'un joueur ? Dans les trois sports présentés ici, les résultats sont obtenus sur la base des informations suivantes : le résultat, le classement et le type de compétition (le modèle prédictif est issu d'une loi de Bernoulli qui ne dépend que du classement). L'un des intérêts de l'attribution d'un classement est de pouvoir "prédire" le résultat d'une confrontation entre deux joueurs. On peut simplement tenter de prédire le résultat d'une confrontation entre deux joueurs qui se sont déjà rencontrés auparavant.Il suffit pour cela de comparer le nombre de victoires de chaque joueur l'un contre l'autre. Le résultat (ou grandeur aléatoire issue du match) est dépendant du niveau de jeu des joueurs qui s'affrontent (paramètres inconnus avant le match). Il y a cependant des limites à cette méthode simple telles que l'évolution du niveau des joueurs depuis leur dernière rencontre ou encore un nombre de rencontres trop faibles pour pouvoir déterminer précisément les probabilités de victoire de chacun. Aussi, si deux joueurs ne se sont jamais rencontrés, il sera impossible avec cette méthode de prévoir le résultat. Il devient alors indispensable d'évaluer le niveau des joueurs au moyen d'un classement. On pourra prévoir l'issue du résultat à l'aide de probabilité de gain de la rencontre, sans même que les joueurs ne se soient déjà affrontés. Ce classement moyen est le plus souvent représenté sous une valeur numérique, attribuée à chaque joueur, reflétant le niveau des joueurs en fonction de leurs résultats passés. Les joueurs sont alors classés sous forme de liste dans l'ordre décroissant de leur nombre de points et peuvent ainsi comparer leur niveau de jeu sans jamais s'être rencontrés. La difficulté est de déterminer une méthode de classement juste et représentative du niveau de chaque joueur.

C'est le physicien américain Arpad Elo, joueur d'échecs confirmé, insatisfait du système de classement aux échecs dans les années 1950, qui crée un système de classement permettant de calculer l'espérance de gain d'un joueur contre un autre en introduisant une loi logistique dans le calcul du classement. Ce système de classement (Elo) permettant la prévision des résultats de matchs, apporte, à l'époque, une grande nouveauté dans le monde des jeux et du sport et représente la base des systèmes de classement modernes. La fonction ELO est $f(x) = 1/(1+10^{-x/400})$ où x est l'écart de classement entre les deux joueurs. Exemple : aux échecs, si un joueur amateur classé 1400 affronte un très bon joueur classé 1700, la probabilité de gain de A est de $f(-300) = 0.15$. Lorsqu'un joueur remporte une partie, son classement est crédité d'un certain nombre de points. Il faut alors définir un barème de gain et de perte de points en fonction des résultats obtenus par les joueurs. On parle alors de méthode de classement "incrémentale". Le principe est d'augmenter ou de diminuer son classement, selon que le joueur gagne ou perd une partie, d'une quantité dépendante de la différence de classement des deux joueurs. En sport, la FIFA a adopté ce classement pour classer les équipes de football féminines. Il existe aussi un site internet (eloratings.net) qui propose un classement mondial, et pour chaque continent, des équipes nationales de football. Dans ce papier, nous revenons sur un aspect saillant de la procédure de classement pour chacun des trois sports. Pour le tennis de table (2), nous avons cherché à définir une règle de handicap de jeu. Au tennis (3), nous revenons sur l'évolution de la procédure de classement via l'analyse de l'influence du nombre de matches disputés dans la saison. Enfin, au golf (4), nous tentons de mettre en correspondance le handicap de jeu avec l'espérance du nombre de coups joués sur un parcours "type".

TABLEAU 1. *Comptage des points FFTT*

Ecart de points	VN	DN	VA	DA
0-24	6	-5	6	-5
25-49	5,5	-4,5	7	-6
50-99	5	-4	8	-7
100-149	4	-3	10	-8
150-199	3	-2	13	-10
200-299	2	-1	17	-12,5
300-399	1	-0,5	22	-16
400-499	0,5	0	28	-20
500 et plus	0	0	40	-29

2. Tennis de Table : Construire une règle de handicap de jeu

2.1. Contexte

En compétition sportive, la mise en place d'un handicap de jeu est un problème souvent délicat. Certains sports comme le tennis ou le golf ont réglé la question en définissant une nomenclature de classement basée sur cette notion de handicap. Au tennis de table, les intitulés des classements, aujourd'hui comme hier, ne traduisent aucune notion de handicap. Pour autant, de nombreux clubs de tennis de table organisent des tournois basés sur des handicaps aussi divers que variés. Nous proposons ici une approche probabiliste permettant de définir, une règle de handicap simple, juste et équilibrée.

2.2. Probabilité de gagner la partie

Au tennis de table, chaque partie en compétition officielle participe au calcul du classement. En fonction de l'issue de la partie et de l'écart de points de classement entre les 2 joueurs, le vainqueur augmente son nombre de points et le perdant le diminue selon un tableau défini par la FFTT. Au tennis de table, les classements débutent à 500 points et vont au delà de 3000 points. Le Tableau 1 nous montre l'échange de points qui s'opère à la suite d'un match entre deux joueurs de classements différents. Par exemple, si un joueur classé 1400 points gagne contre un joueur classé 1230 points, selon le Tableau 1 il aura remporté les 3 points supplémentaires d'une "victoire normale (VN)" (son classement sera alors de 1403 points) alors que son adversaire en aura perdu 2 ("défaite normale (DN)"). Si c'est le joueur classé 1230 qui l'emporte, il ajoutera les 13 points d'une "victoire anormale (VA)" à son classement alors que le joueur classé 1400 points perdra les 10 points correspondant à une "défaite anormale (DA)".

En théorie des jeux (Binmore, 1999), on estime qu'un jeu est équilibré si l'espérance de gain est nulle pour les deux joueurs ("jeux à somme nulle"). Plus précisément, lorsque deux joueurs A

TABLEAU 2. *Probabilités de victoire et défaite (basée sur un jeu à somme nulle)*

Ecart de points	Probabilié VN	Probabilité DN	Probabilité VA	Probabilité DA
0-24	0.455	0.545	0.455	0.545
25-49	0.522	0.609	0.391	0.478
50-99	0.583	0.667	0.333	0.417
100-149	0.667	0.769	0.231	0.333
150-199	0.769	0.867	0.133	0.231
200-299	0.862	0.944	0.056	0.138
300-399	0.941	0.978	0.022	0.059
400-499	0,976	1	0	0.024
500 et plus	1	1	0	0

et B se rencontrent et que la probabilité que A gagne la partie est notée P_A (avec $P_A > 1/2$), alors le jeu est à somme nulle si :

$$P_A K_A + (1 - P_A) K_B = 0 \tag{1}$$

Donc

$$P_A = \frac{K_B}{K_A - K_B} \tag{2}$$

De ce principe, on est en mesure de calculer pour tout K_A et K_B et donc pour tout écart de classement entre deux joueurs, la probabilité que le mieux classé sorte vainqueur (P_A). Le Tableau 2 est le résultat de l'utilisation du Tableau 1 (tableau de la FFTT) pour calculer les probabilités de victoire et de défaite en fonction de l'écart de classement.

Comme on peut le voir dans le Tableau 1, l'attribution actuelle des points n'est pas la solution d'un jeu à somme nulle. En effet, à chaque match disputé, on observe la création de points (il est donné plus de points au vainqueur qu'il en est retiré au perdant). Cette méthode provoque un phénomène d'inflation du classement moyen qui a nécessité la mise en place de la "dérive". Ce nombre est une variable d'ajustement semestrielle qui a pour but de stabiliser le classement moyen des joueurs. Aussi, en fin de semestre, le classement de tout joueur est amputé de la dérive qui est souvent de quelques points. La dérive correspond à la différence entre la moyenne des classement en fin de deuxième phase (semestre 2) et la moyenne des classements en fin de première phase (semestre 1). Normalement (jeux à somme nulle), pour un écart de classement nul, les probabilités de victoires normales et de défaites normales devraient être égales. De même pour les probabilités de victoires anormales et de défaites anormales. Ce n'est pas exactement le cas. En particulier, pour un écart de point entre 0 et 24, nous devrions avoir des probabilités toutes égales à 1/2. L'attribution actuelle des points n'est donc pas la solution d'un jeu à somme nulle. Néanmoins, ces calculs produisent un encadrement de chaque probabilité. Il est alors naturel de prendre comme probabilité, le milieu de chaque intervalle (Figure 2).

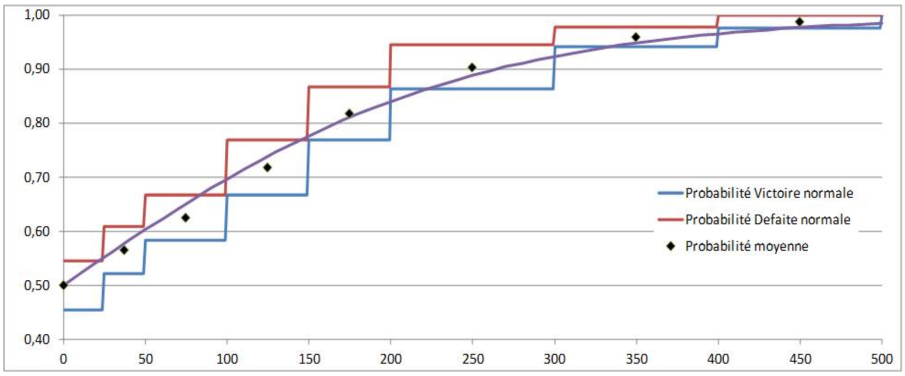

FIGURE 2. *Probabilités de victoire moyennes lissées par le modèle de Bradley-Terry (1952)*

TABLEAU 3. *Probabilités de victoire moyennes lissées par le modèle de Bradley-Terry (1952)*

Ecart de points	0	80	160	240	320	400	480
Probabilité de gagner la partie	0.5	0.66	0.79	0.88	0.93	0.97	0.98

2.3. Probabilité de gagner le point

Au tennis de table, une partie "classique" se joue au meilleur des 5 manches de 11 points. Il est ici possible de mettre en relation la probabilité de victoire d'une partie à handicap $P(p;H)$ en fonction de la probabilité p de victoire du point (voir Sarfati and Fegyvères 2000 pour le tennis sans handicap de jeu).

$$P(p,H) = P_{GM}(p,H)^3 + 3P_{GM}(p,H)^3(1-P_{GM}(p,H)) + 6P_{GM}(p,H)^3(1-P_{GM}(p,H))^2$$

où $P_{GM}(p,H)$ est la probabilité de gagner une manche avec le handicap H.

$$P_{GM}(p,H) = \sum_{i=0}^{9-H} C_{10+i}^{i} p^{11}(1-p)^i + C_{20-H}^{10-H} \sum_{i=10}^{\infty} 2^{i-10} p^{i+2}(1-p)^{i-H}$$

Dans ces équations, $H = 0,-1,-2$, etc. Cela signifie que le joueur le mieux classé commencera chaque manche, non pas à 0 mais, à -1 ou -2, etc, en fonction de sa probabilité a priori de victoire de la partie (sans handicap).
Si $H = 1,2$, etc, cela signifie que le joueur le moins bien classé commence chaque manche non pas à 0 mais à 1 ou 2, etc, en fonction de sa probabilité a priori de victoire de la partie (sans handicap).

2.4. Etapes de construction du handicap de jeu

1. On estime $P(p,0)$ pour chaque écart de classement définit dans le tableau d'attribution des points de la FFTT.

2. Par résolution des équations (1) et (2) pour $H = 0$, on estime la probabilité, p, de gagner un point en fonction de la probabilité de gagner la partie sans handicap $P(p,0)$.
3. Par résolution des équations (1) et (2) pour $H = -1, -2$,etc, on estime la probabilité de gagner le point, p_H, pour un handicap H qui conduit à une probabilité de gagner la partie $P(p_H,H) = \frac{1}{2}$. Ensuite, par calcul direct des équations (1) et (2) pour $H = 0$, on estime alors la probabilité de gagner la partie sans handicap $P(p_H,0)$.
4. En mettant en correspondance $P(p,0)$ et $P(p_H,0)$, on obtient les écarts de points de classement équivalents à chaque handicap H.

2.5. *Résultats*

La Figure 3 montre que la relation entre l'écart de classement et le "juste" handicap est linéaire. On est alors en mesure de proposer un handicap de jeu négatif très simple (Tableau 4).

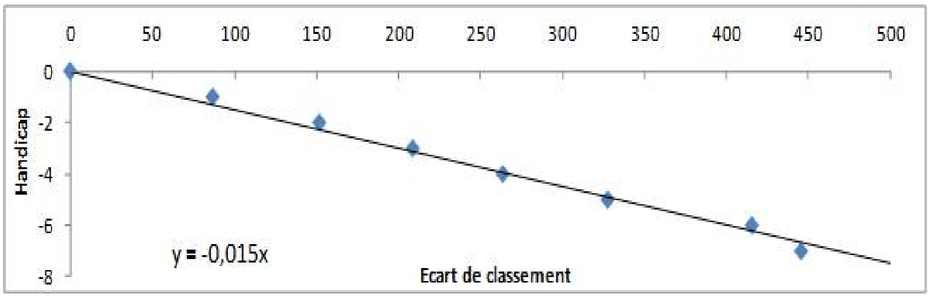

FIGURE 3. *Relation entre écart de classement et handicap de jeu négatif*

Par exemple, un 1430 qui rencontre un 1300 commencera chaque manche à -2. Ce handicap négatif se poursuit théoriquement à -8, -9, etc, tous les 66 points d'écart. Donc, en théorie, chaque manche pourra être très longue par rapport à la même manche sans handicap. Si le joueur le mieux classé démarre la manche avec 0 point, il devra gagner 11 points pour la remporter. Si en revanche il démarre la manche avec -4 points, il devra gagner 15 points pour arriver à 11 et remporter la manche. Le joueur le moins bien classé commencera toujours la manche à 0 points et devra gagner 11 points pour la remporter. On voit bien ici que cette règle augmente le nombre de points que doivent jouer les deux adversaires et par conséquent augmente la durée du match (sachant que le joueur le mieux classé peut très bien commencer à -12 points par exemple). L'inconvénient est donc majeur.
Avec la règle de handicap positif, le joueur le mieux classé commencera toujours la manche à 0 point et devra donc marquer 11 points pour la remporter. Le joueur le moins bien classé commencera la manche à +1, +2, etc, et devra marquer moins de 11 points pour sortir vainqueur. On voit ici, que la règle de handicap de jeu négatif augmente le nombre de coups à jouer dans une manche et donc sa durée, alors que la règle de handicap de jeu positif propose l'effet inverse. Une règle de handicap positif nous paraît donc plus raisonnable. Le temps de chaque manche

TABLEAU 4. *Règle de handicap négatif*

Ecart de classement	0-33	34-100	101-166	167-233	234-300	301-366	367-433	434-500
H du mieux classé	0	-1	-2	-3	-4	-5	-6	-7

TABLEAU 5. *Règle de handicap positif*

Ecart de classement	0-50	51-150	151-250	251-350	351-450	451-550	551-650	651 et +
H du moins bien classé	0	1	2	3	4	5	6	7

est théoriquement le même avec le handicap positif et sans celui-ci. En effet, le joueur le mieux classé a le même contrat à réaliser (marquer 11 points) avec ou sans handicap pour remporter la manche, et le joueur le moins bien classé fournira le même effort pour remporter 3 points (si son espérance de gain de points dans une manche est de 3) qu'il démarre la manche à 0 ou 5 points par exemple. A ce moment là, seule la probabilité de victoire de la manche changera pour les deux joueurs. Pour un système de handicap positif, l'analyse montre que la relation est également linéaire (Figure 4). On est alors en mesure de proposer un handicap de jeu positif également très simple (Tableau 5).

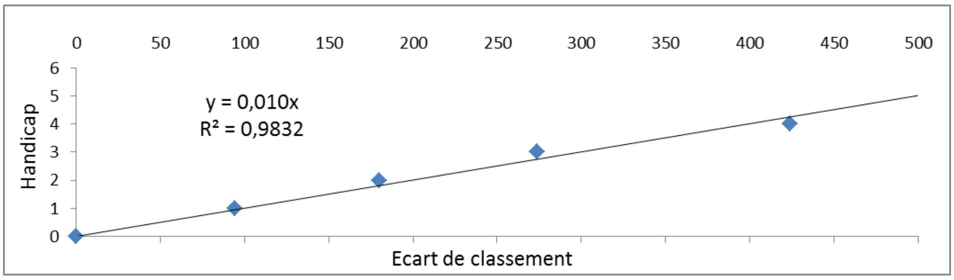

FIGURE 4. *Relation entre écart de classement et handicap de jeu positif*

Par exemple, un 1400 qui rencontre un 1520 commencera chaque manche à +1. Ici il est nécessaire que ce handicap n'augmente pas indéfiniment. En effet, un handicap de +11 donnerait la partie sans jouer. Nous considérons alors que +7 est le handicap maximum acceptable par les joueurs.

2.6. *Conclusion*

Mettre en oeuvre un principe de match par handicap ne doit pas être une opération subjective issue d'un processus d'essai-erreur. C'est une conséquence directe de la procédure de classement elle-même. Une bonne règle de handicap doit "remettre les compteurs à 0". Les deux règles de handicap proposées sont équivalentes et basées sur une probabilité de victoire égale pour les deux joueurs, quelque soit leur différence de classement. On peut se poser la question de l'intérêt pour le joueur le mieux classé de participer à ce type de tournois. En effet, la remise des "compteurs à zéro" est à l'avantage du joueur le plus faible. Afin de conserver un avantage, plus symbolique que significatif, nous proposons de conserver un très léger avantage au joueur le mieux classé. Enfin, nous préconisons le handicap positif car il a le grand avantage de ne pas augmenter l'espérance

TABLEAU 6. *"Avantages" de jeu en fonction de l'écart de classement*

Ecart de classement	0-99	100-199	200-299	300-399	400-499	500-599	600-699	700 et +
H du moins bien classé	0	1	2	3	4	5	6	7

du temps de jeu de chaque manche. Aussi, la règle "d'avantage" de jeux que nous proposons est la suivante Tableau 6.

Par exemple, si l'écart de classement entre les deux joueurs est de 215 points, le joueur le moins bien classé commencera chaque manche avec 2 points.
Si les matchs disputés avec la règle de handicap devait entrer dans le calcul du classement des joueurs, on utiliserait alors la première ligne du Tableau de la FFTT (1) pour l'échange de points à l'issu du match.

3. Tennis : La procédure de classement est-elle sensible au nombre de matches disputés durant la saison ?

3.1. Contexte : Evolution du classement au Tennis depuis 1977

Cette étude présente une approche généralisée pour l'analyse de la procédure de classement au tennis depuis son informatisation en 1977. Une précédente étude (Paris and Gerville-Réache, 2005) avait mis en évidence une croissance de la fonction médiane du bilan avec le nombre de matches disputés. Concentrée sur l'année 2005, cette étude utilisait déjà la simulation probabiliste mais également le palmarès réel des joueurs. Pour une analyse longitudinale, il est nécessaire de s'abstraire de la réalité du classement des joueurs qui, par définition, est dépendant de la procédure de classement elle-même. La nouvelle modélisation probabiliste permet ainsi de visualiser l'évolution des exigences de la procédure de classement de 1977 à nos jours (Gerville-Réache and Paris, 2009)).

3.2. Procédure de classement

L'équation du bilan d'un joueur (i) peut se formaliser sous la forme suivante (voir Annuaire officiel du classement, Fédération Française de Tennis, 1977-2006).

$$\max_{a<A,b<B,c<C,d<D,e<E,f<F,} B_i = C_i + (150a + 100b + 50c + 30d + 20e + 15f)$$

Ce bilan (B_i) est comparé aux normes fédérales de maintien et de montée pour déterminer le classement futur du joueur. C_i est ici le capital de départ, A,B,C,D,E et F sont les nombres de victoires aux différents échelons. $V = a+b+c+d+e+f$ est le nombre de victoires prises effectivement en compte dans le bilan. Le nombre de victoires prises en compte dans le bilan dépend du nombre total de victoires (v), du nombre de défaites significatives à échelon égal (E), à un échelon inférieur (I) et à deux échelons inférieurs (G), et se traduit par : $V = v - E - 2I - 5G$

3.3. Approche par simulation probabiliste

Simuler la saison d'un joueur de tennis dans le but d'analyser la procédure de classement, nécessite d'identifier les facteurs influançant le bilan. Clairement, le calcul du bilan est déterminé par : le nombre de matches disputés dans la saison, et pour chaque match, le classement de l'adversaire et l'issue du match (victoire ou défaite).
Le principe de la simulation est donc de générer une succession de matches. Le joueur (simulé) va donc disputer un premier match. Contre qui ? Sort-il vainqueur ? Ces deux questions nécessitent alors de probabiliser le niveau de l'adversaire puis l'issue du match.

3.3.1. Probabilités de victoire

A l'origine du tennis en France, les classements étaient l'expression d'un handicap laissé par le joueur le mieux classé contre le joueur le moins bien classé. Par exemple, lorsque le joueur dit de référence, le 0, affrontait un 2/6, le handicap était d'un point, deux jeux sur six. Ce handicap laissé était censé rééquilibrer le match. Les points laissés traduisaient donc une mesure de l'écart de niveau théorique entre les joueurs. Fort de ce principe, il est alors possible, par inversion, d'obtenir, par simulation, les probabilités de victoire (Figure 5) d'un match entre un joueur de classement H (le classement de référence) et un joueur mieux classé (par exemple $H-2$) ou un joueur moins bien classé (par exemple $H+1$).

Selon Stewart (1992) on peut obtenir la probabilité de gagner le match d'après la probabilité p de gagner un point avec $q = 1-p$.

1. Probabilité P_1 de gain d'un jeu :

$$p_1 = p^4 + 4p^4q + \frac{10p^4q}{1-2pq}$$

2. Probabilité p_2 de gain d'un jeu décisif :

$$p_2 = p^7 + 7p^7q + 28p^7q^2 + 84p^7q^3 + 210p^7q^4 + \frac{462p^7q^5}{1-2pq}$$

3. Probabilité p_3 de gain d'un set avec jeu décisif $(q_1 = 1-p_1)$:

$$p_3 = p_1^6 + 6p_1^6q_1 + 21p_1^6q_1^2 + 56p_1^6q_1^3 + 126p_1^6q_1^4 + 252p_1^6q_1^5 + 504p_1^6q_1^6p_2$$

4. Probabilité p_4 de gain d'un match $(q_3 = 1-p_3)$:

$$p_4 = p_3^3 + 3p_3^3q_3 + 3p_3^3q_3^2$$

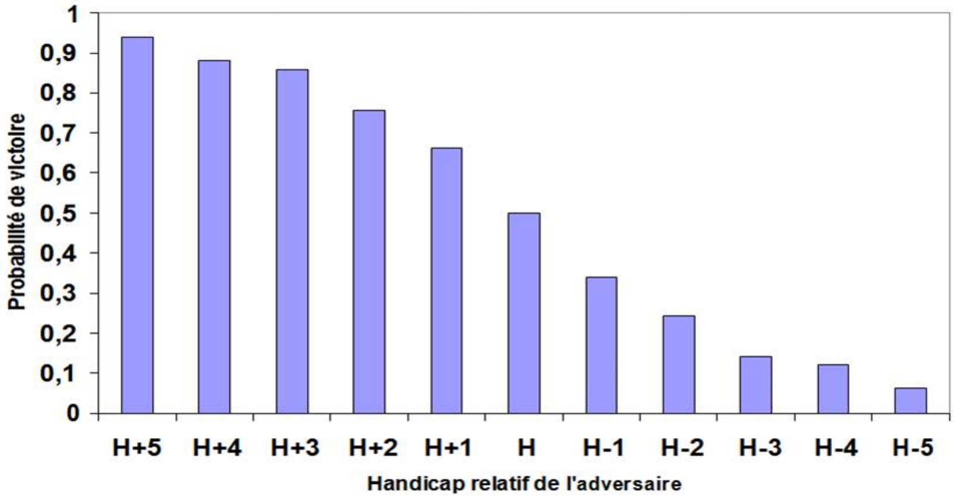

FIGURE 5. *Probabilité de victoire en fonction du Handicap (classement)*

3.3.2. Probabilités de rencontre

A partir du modèle de probabilité de victoire, on va pouvoir définir un modèle de probabilité de rencontre. Pour chaque match joué, il faut associer une probabilité que celui-ci soit joué contre un adversaire d'un échelon donné. H sera l'échelon de référence et H_i celui de l'adversaire. Les probabilités de rencontre sont alors obtenues par simulations de matches sur des tournois complets (où tous les échelons du classement sont inscrits), avec compression (c'est à dire avec la possibilité de jouer à classement égal), en utilisant les probabilités de victoire obtenues précédemment (Figure 5). On simule un match de tennis opposant un joueur X à un joueur Y. La simulation commence par le tirage pseudo aléatoire d'un nombre compris entre 0 et 1 suivant une loi uniforme. Quand ce nombre est inférieur à la probabilité de X de battre Y, X est vainqueur. Dans le cas contraire, Y est vainqueur. Lorsque la simulation d'un match est terminée, on peut simuler le tournoi entier. On est alors en mesure de simuler le tournoi 10 000 fois afin d'estimer les probabilités de rencontre (Figure 6).

3.3.3. Génération d'un match, d'une saison, de 1000 saisons...

Les systèmes de probabilités étant définis (pour un joueur dont le classement reflète son niveau de jeu réel), on sélectionne au hasard un adversaire selon les probabilités de rencontre puis, on sélectionne au hasard l'issue du match selon les probabilités de victoire. Un match a donc été simulé. En répétant 10 fois l'opération, on peut calculer le bilan d'un joueur qui aurait disputé 10 matches dans la saison. En répétant 1000 fois la simulation des 10 matches, on obtient alors une approximation très satisfaisante de la distribution des bilans d'une saison à 10 matches. En faisant varier le nombre de matches disputés dans la saison, on est en mesure d'estimer le nombre de matches permettant d'avoir, par exemple, une chance sur deux de monter, se maintenir ou descendre dans le classement. En réalisant cela pour tous les classements, on obtient un outil d'analyse de la procédure de classement aussi précis que l'on veut.

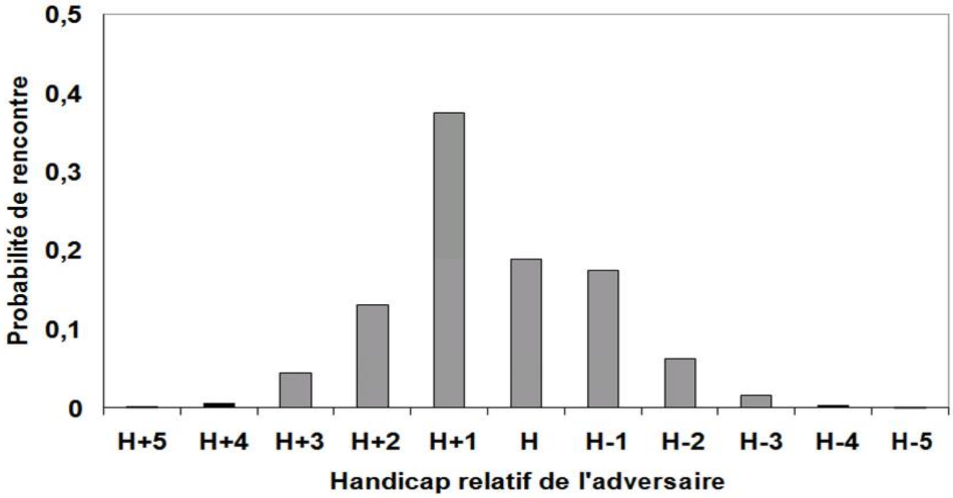

FIGURE 6. *Probabilité de rencontre*

3.4. Résultats

Les Figures suivantes 7 et 8 montrent que les exigences du classement ont eu une évolution relativement irrégulière et inégale. On peut voir, par exemple en 2006, qu'un joueur classé 15 a une chance sur deux de monter dans le classement (sans changer de niveau de jeu) à partir de 40 matches disputés. Il en fallait au moins 150 pour un -2/6. En 1995, il en fallait 20 pour le joueur classé 15 et seulement 45 pour le -2/6.

3.5. Conclusion

La procédure de classement est une équation stochastique composée de 6 variables aléatoires (A, B, C, D, E et F) de lois inconnues. Une étude mathématique est donc d'une complexité difficilement surmontable. L'approche par simulation purement probabiliste est, sans doute, l'unique voie pour une analyse concrète et objective. Ces évolutions, pour le moins inattendues, mériteraient, sans doute, un décryptage par la FFT. En 2013, la FFT a souhaité opérer quelques modifications à la procédure de classement : la suppression du capital de départ (afin de simplifier le calcul du classement), la suppression des bonus aux vainqueurs de tournois (afin que le classement reflète une vérité sportive réellement acquise sur le terrain), une pénalisation plus importante pour les joueurs ayant un mauvais bilan et un réajustement des barèmes de gain et de perte de points.

4. Autour du classement au Golf : handicap et probabilités de performances

4.1. Construction du classement au golf

Le golf est l'un des seuls sports dans lequel vous pouvez jouer contre un joueur de niveau différent sans vous ennuyer, grâce à un système de coups rendus qui dépend de votre index (votre classement). Cette formule de jeu se rapproche de la règle de handicap de jeu positif développée

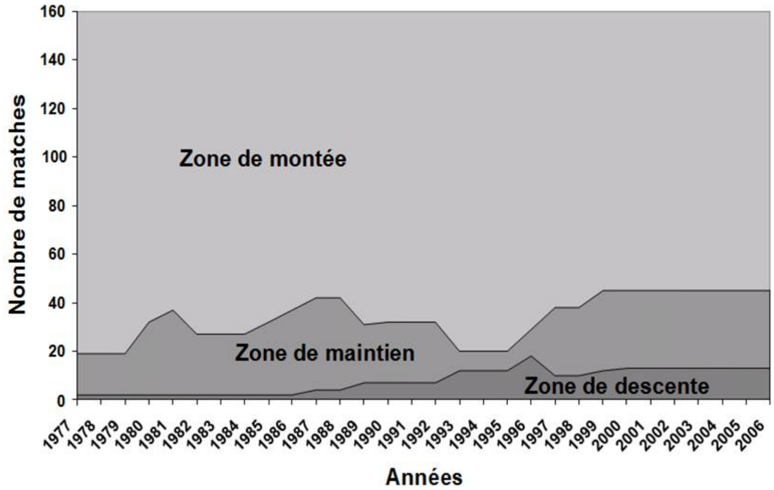

FIGURE 7. *Evolution des exigences médianes du classement pour un joueur classé 15 : 1977 à 2006*

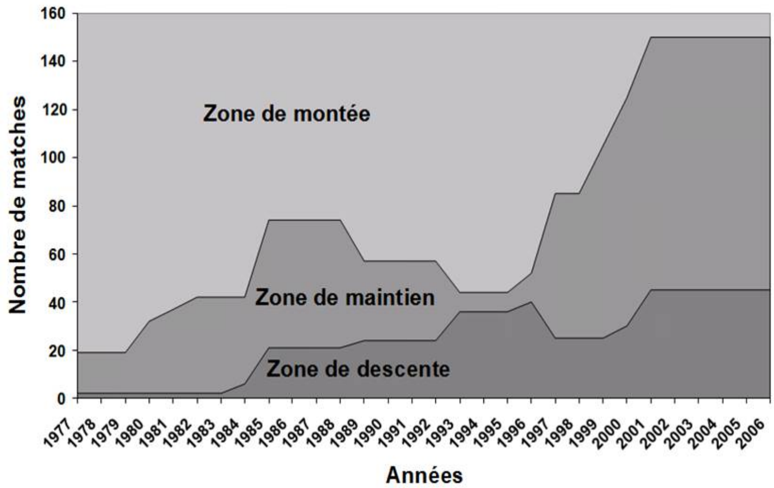

FIGURE 8. *Evolution des exigences médianes du classement pour un joueur classé -2/6 : 1977 à 2006*

pour le tennis de table.

Pour comprendre le calcul du classement (index) au golf, il faut prendre en compte 3 paramètres :

- le slope mesure la difficulté du parcours : plus le parcours est compliqué plus le slope sera élevé et plus le joueur aura des "coups bonus" ou "rendus" en fonction de son index ;
- le Stableford est une méthode de comptage de points d'une partie de golf très utilisée de nos jours. Le score de référence est de 36 points car le par (ou nombre de coups théorique fixé pour un trou) vaut 2 points sachant qu'un parcours contient 18 trous. Il en existe deux types : le Stableford net permet de ne pas trop pénaliser un joueur en difficulté sur un trou. Au début d'une partie un joueur peut, selon son index, se voir rendre des coups en fonction de la difficulté du parcours (handicap), cela permet d'équilibrer les scores entre joueurs de niveaux différents (il permet aussi de déterminer quel a été le meilleur joueur de la compétition par rapport à son handicap de jeu) ;

TABLEAU 7. *A : Catégories B : Bornes d'index C : Coefficient de descente pour une perf par point joué en dessous D : Coefficient de remontée pour une contre performance E : Marge de contre performance sans remontée (point stableford)*

A	B		C	D	E
	de	à			
1	0 jusqu'à	4.4	0.1	0.1	1 point
2	4.5	11.4	0.2	0.1	2 points
3	11.5	18.4	0.2	0.1	3 points
4	18.5	26.4	0.4	0.1	4 points
5	26.5	53.5	0.5	0.2	5 points

le Stableford brut (basé uniquement sur le nombre de coups réalisés sur le parcours) permet quant à lui de déterminer quel a été le meilleur joueur de la compétition ;

- Le barème de baisse et de remontée de l'index : au golf, il est assez facile de descendre (améliorer son index/handicap) et plus difficile de remonter. Les joueurs sont classés dans 5 catégories différentes, selon leur classement, auxquelles sont associés un coefficient de descente et un coefficient de remontée. Les coefficients de descente sont majoritairement plus élevés que les coefficients de remontée, ce qui semble favoriser l'amélioration de l'index. A cela s'ajoute une "zone tampon" qui protège l'index des joueurs en cas de petite contre-performance et qui est ajustée sur la base des performances de tous les joueurs (Tableau 7).

Par exemple, un joueur d'index 25, joue sur un parcours qui lui rend 24 coups. Le par du parcours est de 72 points. Le par du joueur (dépendant de son index et de la difficulté du parcours) est de : $72+24=96$. Les coups rendus sont répartis un par un selon l'ordre de difficulté des trous. En réalisant un score Stableford net de 40, il a joué quatre coups en dessous de "son" par (soit 92) et réalise une performance. Il verra donc son index diminuer de quatre fois le coefficient de descente pour son index (ici 0.4) soit $4*0.4=1.6$ points. Le nouvel index est donc $25-1.6=23.4$.

4.2. Le Handicap

Extrait du règlement European Golf Association (EGA) Handicap System Amended Edition, 1 January 2012 :
"Tous les handicaps sont liés, et découlent, des résultats historiques et actuels du joueur. Le système n'est pas conçu pour qu'un joueur joue à son handicap, ou en dessous, aussi fréquemment qu'il y joue au-dessus. Par exemple, il est probable qu'un joueur correctement "handicapé" de catégorie 1 joue à son handicap, ou en dessous, 35% du temps ; alors qu'un joueur correctement "handicapé" de catégorie 4 atteindra son handicap ou un meilleur résultat seulement 10% du temps. Un joueur en progression jouera en dessous de son handicap plus fréquemment qu'un joueur au handicap "réel" identique. Cela continuera jusqu'à ce qu'il atteigne son handicap réel."
Cet extrait de la règle de handicap au golf et l'objectif qui en découle, nous permettent de mettre en place une distribution de probabilités des coups "gagnés" et "perdus" sur un parcours type, par

rapport au handicap du joueur, et cela pour les 5 catégories.

4.3. Proposition de modélisation

Notre objectif ici est de modéliser les pourcentages de joueurs qui réaliseront un nombre de coups inférieur ou égal à leur handicap pour chaque catégorie, sur un parcours type.
La première idée serait d'effectuer une expérimentation sur le terrain afin d'en ressortir des statistiques de coups gagnés et perdus pour des joueurs de toutes les catégories de l'index. Néanmoins, nous ne sommes pas en mesure de connaître le niveau réel des joueurs. Ainsi, l'approche théorique nous semble être la meilleure option d'analyse.

Nous partons de l'information théorique (du règlement précédent) pour la modélisation :
- Environ 10% des golfeurs de catégorie 4 réaliseront un nombre de coups inférieur ou égal à leur handicap.
- Environ 35% des golfeurs de catégorie 1 réaliseront un nombre de coups inférieur ou égal à leur handicap.

Nous devons maintenant construire les probabilités de réalisation du nombre de coups inférieur ou égal au handicap des joueurs des trois autres catégories (2, 3, 5). Pour cela, partant de deux points (pourcentages pour la catégorie 1 et 4) nous procédons à un ajustement de la droite avec une fonction de type puissance (Figure 9).

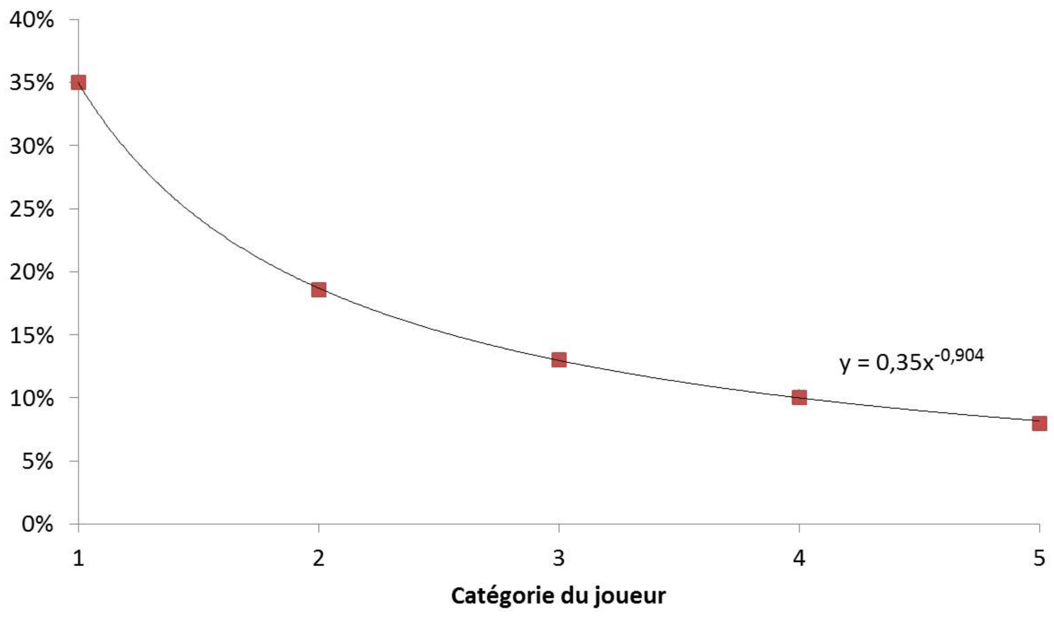

FIGURE 9. *Pourcentage de joueurs qui joueront un résultat inférieur ou égal à leur Handicap de jeu*

De cet ajustement découlent les pourcentages théoriques de joueurs qui réaliseront un nombre de coups inférieur ou égal à leur handicap pour chacune des catégories manquantes (Tableau 8) :

TABLEAU 8. *Nombre de coups inférieur ou égal au Handicap par catégorie*

Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
35%	19%	13%	10%	8%

A présent, pour chaque catégorie, il nous faut construire une distribution des probabilités de score qui vérifie des propriétés de croissance, décroissance et de plateau relatif à la zone tampon correspondante (Figure 10).

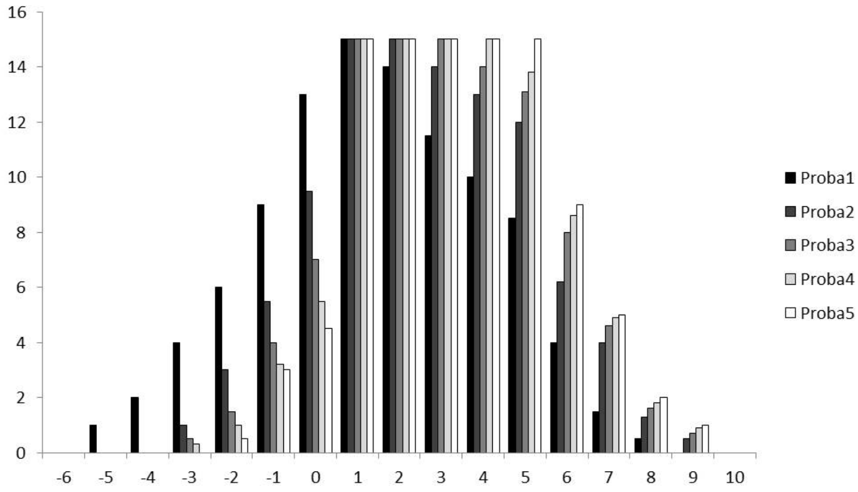

FIGURE 10. *Distribution des probabilités de coups gagnés et perdus*

Soit P_{ij}, la probabilité qu'un joueur de catégorie $i (i = 1$ à $5)$, réalise un score net $j (j = -6$ à $10)$. Soit Z_i la zone tampon de la catégorie i.
On propose que :

$\forall i, \forall j < 1, P_{i,j+1} > P_{ij}$

$\forall i, \forall j \in Z_i, P_{ij} = P_0$ (constante et identique).

$\forall i, \forall j$ au delà de la zone tampon, $P_{i,j+1} < P_{ij}$

Pour chaque $j < Z_i$, quel que soit $i' > i, P_{i'j} < P_{ij}$

Pour chaque $j > Z_i$, quel que soit $i' > i, P_{i'j} > P_{ij}$, $\sum_j P_{ij} = 1; \forall i$

Pour chaque catégorie i, $\sum_j P_{ij} = q_i; \forall i$, q_i suivant le Tableau 9 :

Ces hypothèses de travail permettent de calculer les espérances en nombres de coup par catégorie (Figure 11).

TABLEAU 9. *Probabilités de jouer un nombre de coups inférieur ou égal au Handicap par catégorie*

Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
0.35	0.186	0.13	0.1	0.08

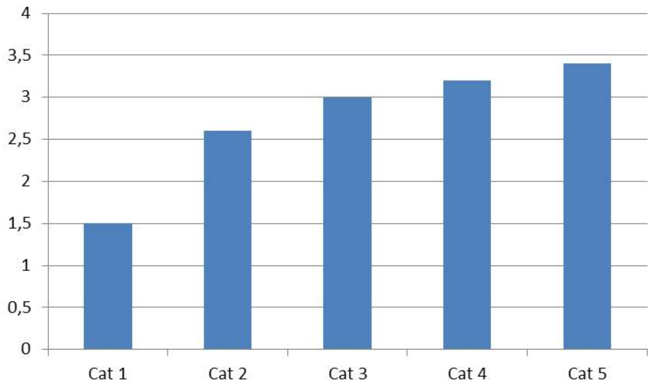

FIGURE 11. *Espérances en nombres de coup par catégorie*

Par exemple un joueur de catégorie 3, jouera en moyenne 3 coups au-dessus de son handicap de jeu. NB : Dans une compétition de golf, le résultat "moyen" des meilleurs joueurs est pris en compte avec un ajout ou un retrait de point. Aussi, il est possible que bien que votre score stableford soit par exemple de 36, vous obteniez une amélioration de votre index car les meilleurs joueurs auraient réalisé de mauvaises performances. Les scores des meilleurs indiquant la difficulté des conditions de jeu le jour de la compétition.

4.4. Conclusion

Les joueurs de golf ont de nombreuses maximes. L'une d'elle dit : "Au golf, il y a les bons jours et les mauvaises années". Au regard de notre analyse, cette maxime semble particulièrement fondée. En effet, même si la plupart des joueurs ignorent que la procédure de classement au golf fait qu'il est bien plus difficile de réaliser une performance qu'une contre performance, ceux-ci le vivent au quotidien. Un joueur (correctement classé) de 5ème ou 4ème série (les séries qui regroupent la plupart des joueurs), jouera moins d'une fois sur dix à son handicap de jeu (ou moins encore). En moyenne, il jouera plus de 3 coups au dessus de son handicap. Cette réalité de l'index au golf en fait une méthode de classement atypique. En effet, au tennis, comme au tennis de table, le principe de base est que vous avez une chance sur deux de gagner contre un joueur de même niveau que vous. Au golf, vous jouez contre un "vous-même" qui est en réalité plus fort que vous de quelques coups. Pas si facile de gagner contre ce "vous-même" !

5. Conclusion générale

Chaque sport possède une procédure de classement spécifique. Afin d'en comprendre les propriétés, il est nécessaire de mettre en place une approche systématique et rigoureuse. Le parti pris de

ce papier a été de considérer le niveau de jeu "moyen" comme un paramètre inconnu. Au cours d'un match donné, le ou les joueurs obtiennent un niveau aléatoire qui dépend de ce niveau de jeu moyen. La procédure de classement est alors la procédure statistique qui est censée estimer ce niveau de jeu (moyen). Le classement qui en découle, et qui correspond donc au niveau de jeu moyen des joueurs, évolue au cours de leur carrière sportive et ceci une à plusieurs fois au cours de la saison, voire entre chaque compétition selon le sport. De nombreuses recherches portent sur la construction d'une procédure de classement ayant des propriétés statistiques classiques (sans biais, rapidement convergente...). Cette tâche est rendue délicate par l'atypicité du type de données : la matrice des rencontres de l'ensemble des compétiteurs d'une fédération est essentiellement vide. Notre regard s'est porté, non pas sur les qualités statistiques des procédures de classement, mais sur des caractéristiques spécifiques permettant de développer des stratégies de jeu ou encore des règles de jeu pour les tournois. Aussi, c'est conditionnellement aux procédures de classement que nous avons réfléchi. Au tennis de table, nous avons déduit de la procédure de classement en vigueur une proposition de règle de handicap de jeu. Au tennis, nous avons donné une lecture de l'évolution de la procédure de classement depuis son informatisation en 1977. Enfin, au golf, nous avons tenté d'expliquer une maxime bien connue des joueurs.
Plus généralement, chaque procédure de classement a des propriétés spécifiques dont les raisons sont souvent profondes et liées à la nature du sport associé. Le Tableau 10 résume quelques-unes de ces propriétés.

La procédure de classement est une des "vitrines" d'un sport. Si on la considère comme injuste, ou trop difficile, on se détournera de ce sport pour aller peut-être vers un autre. Si, par exemple, le risque de défaite et son impact sur son classement est important, le joueur hésitera alors à participer aux compétitions. Dans tous les cas, choisir une procédure de classement est un réel casse-tête pour chaque fédération sportive. A la vue des évolutions successives de la procédure de classement au tennis, il semble qu'un accompagnement théorique soit nécessaire.

TABLEAU 10. *Propriétés des procédures de classement : golf, tennis et tennis de table*

Le Classement	Golf	Tennis	Tennis de table
Evolution du classement du joueur	A chaque compétition	Tri-annuelle	Mensuelle
Variabilité de l'impact d'un résultat	Variabilité croissante en fonction du classement (rapport 1 à 5)	Pas de variabilité	Variabilité croissante en fonction de l'importance de la compétition
Influence du nombre de matches dans la saison	A étudier précisément, mais sans doute importante...	Important et dépend du classement	Faible et indépendante du classement
Tolérance aux petites contre performances	Zone tampon qui dépend du classement	Léger tampon via le nombre de victoire	Aucun tampon
Impact d'une contre performance importante	Indépendant de l'ampleur de la contre (Binaire)	Limité à 2 échelon (3 niveaux)	Va jusqu'à un demi classement (9 niveaux)

Références

Astié, A. (1970). Comparaisons par paires et problèmes de classements. estimation et tests statistiques. *Math & Sci. hum.*, 32 :17–44.

Binmore, K. (1999). *Jeux et théorie des jeux*. DeBoeck Université.

Bradley, R. and Terry, M. (1952). Rank analysis of incomplete block designs. *Biometrika*, 39 :324–345.

Brossard, P. (1994). Quel avenir pour le classement ? *Tennis de France*, 481 :70–78.

Champagne, L., Coupet, A., and Gerville-Réache, L. (2012). Construction d'une règle de handicap de jeu au tennis de table : une approche probabiliste. In *44ème Journées de statistique, Bruxelles, Belgique*.

Champagne, L., Coupet, A., and Gerville-Réache, L. (2013). Elaborating a rule for game handicap in table tennis : a probabilistic approach. In *International Journal of Table Tennis Sciences n8*, pages 66–67.

Coulom, R. (2010). Jeux et sports : le problème des classements. *Pour la science*, 393 :20–27.

Coupet, A. and Gerville-Réache, L. (2007). Comparison between table tennis scoring systems (11 and 21 points) by probabilistic simulation. In X.P. Zhang, D. X. and Dong., Y., editors, *People's Sports Publishing House of China*, pages 33–40. The proceedings of the ninth International Table Tennis, Federation sports science congress.

Courcoux, P. and Séménou, M. (1997). Une méthode de segmentation pour l'analyse de données issues de comparaisons par paires. *Revue de statistique appliquée*, 45 (2) :59–69.

Delamarre, G. (1973). L'ordinateur bouleverse le classement. *Tennis de France*, 239 :64–66.

FFTT (2014). *Tableau de comptage des points FFTT*. Accessed : 2014 - 01 - 10.

Gerville-Réache, L. and Paris, N. (2009). Evolution de la méthode de classement au tennis : Approche par simulation probabiliste. In *13ème Congrés International de l'ACAPS*, pages 563–564.

Glickman, M. and Jensen, S. (2003). Adaptive paired comparison design. *JSPI*, 127 :279–293.

Hallinan, S. (2005). Paired comparison models for ranking national soccer. Technical report, Worcester Polytechnic Institute.

His, B. and Burych, D. (1971). Games of two players. *Applied Statistics*, 20 :86–92.

Klaassen, F. and Magnus, J. (2001). Forecasting the winner of a tennis match. *Econometrics*, Center Discussion Paper (Int. r. no. 2001-38) :20.

Magnus, J. and Klaassen, F. (1999). On the advantage of serving first in a tennis set : Four years at wimbledon. *The Statistician*, 48 :247–256.

Paris, N. (2008). *Formalisation algorithmique des classements au tennis. Mise en perspective longitudinale par simulation probabiliste*. PhD thesis, Université Bordeaux Segalen.

Paris, N. and Gerville-Réache, L. (2005). Etude du classement au tennis : Modélisation et analyse statistique par la méthode de monte-carlo. *Math & Sci. hum*, 170 (2) :47–55.

Parlebas, P. (2005). Modélisation dans les jeux et les sports. *Mathématiques & Sciences humaines Mathematics and Social Sciences*, 170 :11–45.

Sarfati, S. and Fegyvères, M. (2000). *Mathématiques : Méthodes, savoir-faire et astuces*. Bréal.

Stewart, I. (1992). *La mathématique des jeux*. Pour la science.