

Stratégies d'arbitrage dans les paris sportifs

La mise en place de stratégies d'arbitrage dans les paris sportifs associe à la fois des considérations probabilistes et analytiques. Nous avons été captivés par cette méthode méconnue du grand public, dont l'approche demeure non triviale, mais très instructive pour pallier aux ajustements des cotes établis par les bookmakers.

Ainsi, cette notion, qui peut se résumer in fine à l'étude de martingales, s'inscrit convenablement dans le thème de l'année puisqu'elle associe le domaine sportif à l'aspect probabiliste que mettent en jeu les situations de paris sportifs.

Ce TIPE fait l'objet d'un travail de groupe.

Liste des membres du groupe :

- *BIER Oscar*

Positionnement thématique (ÉTAPE 2) :

- *INFORMATIQUE (Informatique pratique)*

- *MATHEMATIQUES (Autres)*

Mots-clés (ÉTAPE 2)

Mots-clés (en français)

Paris sportifs

Stratégie d'arbitrage

Processus d'extraction de données

Sous-martingale

Processus stochastique

Mots-clés (en anglais)

Sport bets

Arbitrage strategy

Web scraping

Sub-martingale

Stochastic process

Bibliographie commentée

L'étude de stratégies d'arbitrage occupe une place significative au sein de la recherche dans le domaine statistique. Au cours des dernières décennies, elle a gagné en importance, notamment en raison de ses multiples applications dans le domaine financier, que ce soit en bourse pour investir dans des actifs sans risque, ou plus généralement dans des contextes impliquant des résultats aléatoires tels que les paris sportifs ou encore les jeux d'argent.

Les paris sportifs furent légalisés à la fin du XIXème siècle, et devinrent rapidement populaires durant le XXème siècle. Désormais, ils font l'objet d'une véritable étude approfondie, étant donné leur lien étroit avec la théorie des jeux de hasard mais aussi avec le domaine de la gestion financière. Plusieurs problèmes peuvent être soulevés : existe-t-il une stratégie qui permettrait de gagner à tous les coups ? A défaut, peut-on tout de même minimiser le risque, ou bien le hasard l'emporterait-il toujours ?

Tout événement d'un pari sportif est associé à une cote, qui désigne la retranscription de la probabilité que cet événement se réalise, dont la valeur est fixée par les bookmakers. Cette décision découle d'une analyse approfondie des éléments liés au match, englobant les antécédents de performance des joueurs, les conditions de jeu et leur état de santé. Toutes ces données sont soumises à un traitement à l'aide de modèles, notamment la régression logistique, fréquemment utilisée pour déterminer les liens entre ces diverses variables. Ces relations sont ensuite organisées au moyen d'algorithmes de classification [4]. Toutefois, les cotes peuvent être légèrement modifiées en raison d'une marge que souhaite se laisser le bookmaker, ainsi que d'une anomalie, le Favourite Longshot Bias, stipulant que l'espérance de gain des joueurs favoris est généralement supérieure à celle des outsiders [5].

La martingale est un outil de modélisation bien adapté à la situation. Apparue au cours des années 1930 suite aux travaux de Lévy, Bernstein et Ville [3], elle est définie comme une séquence de variables aléatoires telle que l'espérance conditionnelle demeure constante. Véritable fondement du concept de processus stochastique qui est l'étude des phénomènes aléatoires au cours du temps, la martingale, à première vue simpliste, renferme en réalité toute une théorie complexe, permettant une fine description d'un phénomène aléatoire quelconque. Cet outil a pu servir à modéliser plusieurs problèmes connus tels que l'urne de Polya, le processus de branchement de Galton-Watson ou encore la ruine du joueur [2]. Plusieurs théorèmes de comportement asymptotique, en particulier le théorème d'arrêt de Doob, sont en effet valables pour les martingales. La convergence certaine d'une martingale à valeurs positives en est l'un des fondamentaux. En outre, la question du temps d'arrêt est centrale dans les situations de paris sportifs. Est-il possible d'anticiper la durée d'attente nécessaire pour obtenir un gain spécifique, ou est-ce que cela dépend uniquement du hasard [1] ?

Afin de vérifier expérimentalement la pertinence de la modélisation, le web scraping, ou crawling, est une méthode efficace pour accéder aux cotes des paris sportifs, qui s'appuie sur une extraction de données en temps réel à partir de sites web dédiés [4].

Problématique retenue

Il s'agit d'établir une modélisation et une résolution satisfaisante de la situation d'un pari sportif, afin de mettre en exergue des stratégies permettant d'optimiser son gain.

Objectifs du TIPE du candidat

Je me propose :

- de modéliser une situation de pari sportif comportant deux issues par une martingale,
- d'étudier quelques propriétés asymptotiques (convergence, temps d'arrêt, théorème de Doob),
- d'extraire des données sur des sites de paris sportifs, dans le cas particulier d'un match de tennis.

Références bibliographiques (ÉTAPE 1)

- [1] THIERRY BODINEAU : Modélisation de phénomènes aléatoires : introduction aux chaînes de Markov et aux martingales : *Cours de l'école Polytechnique, pages 123-209, 2023*
- [2] LAURENT SERLET : Pour dompter l'aléatoire, rien ne vaut une bonne martingale : <https://lmbp.uca.fr/~serlet/mart-mea.pdf>, 2014, consulté le 14 Octobre 2023
- [3] VLADIMIR POZDNYAKOV, MARTIN KULLDORFF : Waiting times for patterns and a method of gambling teams : *The American Mathematical Monthly, pages 134-137, February 2006*
- [4] PAUL STEFFEN : Statistical modeling of event probabilities subject on a sports bet: Theory and applications to soccer, tennis and basketball : *Thèse de doctorat, Université de Bordeaux, pages 14-20 & 47-49, 2022*
- [5] CHRISTOPHE BARRAUD : L'Efficiencence informationnelle du marché des paris sportifs : un parallèle avec les marchés boursiers : *Thèse de doctorat, Université Paris Dauphine, pages 61-68, 2012*
- [6] CLÉMENT DE GROC : Collecte orientée sur le Web pour la recherche d'information spécialisée : *Thèse de doctorat, Université Paris-Sud, pages 17-23, 2013*
- [7] STEVEN SKIENA : Calculated bets : computers, gambling, and mathematical modeling to win : *Cambridge University Press, pages 33-34 & 140-147 & 205-208, 2001*

DOT

- [1] : Recherche d'une stratégie de pari gagnante par l'étude du fonctionnement des cotes
- [2] : Identification d'une stratégie satisfaisante: la stratégie d'arbitrage
- [3] : Découverte de l'outil de la martingale qui nous a mené à la lecture des références bibliographiques [1], [2] et [3]
- [4] : Echec d'une première modélisation de la stratégie d'arbitrage avec une martingale
- [5] : Suite à la lecture de la thèse [6], mise en place d'algorithmes de récoltes des données à l'aide des modules "Beautifulsoup" et "Selenium"

[6] : Difficultés rencontrées à cause du manque de données de certains joueurs de tennis. Il s'est donc avéré fructueux d'utiliser un fichier log pour identifier ces erreurs sans réinitialiser l'algorithme de récoltes

[7] : Réussite d'une modélisation satisfaisante avec une sous-martingale: outil plus adapté pour cette stratégie

[8] : Compréhension et application du temps d'arrêt et du théorème de Doob pour l'étude de propriétés asymptotiques