

## LE REMPLACEMENT AU BILLARD CARAMBOLE COMME PROCESSUS DE MARKOV

MATHIEU BOUVILLE,\* *Institute of Materials Research and Engineering, Singapore*

### Résumé

Le remplacement est la pierre angulaire des jeux de séries au billard carambole (billard français) : sur un coup facile le joueur peut marquer tout en s'assurant que le coup suivant sera favorable. Un coup facile est ainsi généralement suivi d'un autre coup facile. La difficulté d'un coup dépend donc du coup précédent. On présente un processus de Markov qui tient compte de ces corrélations. Ce modèle permet d'expliquer les longues séries de coups faciles et les scores élevés qui en résultent. Il permet aussi d'identifier des différences dans la nature des points sur lesquels des joueurs de niveau différent marquent le plus. Les joueurs peuvent utiliser ce modèle *via* la page <http://billiards.mathieu.bouville.name/biMar/>.

Mots clés : billard, sport, processus de Markov, processus de Bernoulli, traduction

*Le présent article est la traduction de « position play in carom billiards as a Markov process » publié dans la revue Journal of Quantitative Analysis in Sports [1]. (Pour les non-spécialistes, « Le français, le suisse, le russe et l'américaine » [2] présente les probabilités bernoulliennes et markoviennes dans le cadre du billard carambole.) La présente traduction est disponible à [http://billiards.mathieu.bouville.name/Bouville-billard\\_Markovien.pdf](http://billiards.mathieu.bouville.name/Bouville-billard_Markovien.pdf). La version anglaise demeure la version de référence.*

### 1. Introduction

Le billard carambole (ou « billard français ») se joue avec trois billes sur une table sans poches (un court glossaire est proposé en annexe). Si la bille du joueur touche les deux autres billes le joueur marque un point et rejoue. Quand il manque c'est au tour de son adversaire de jouer.

En 1880, à Paris, l'américain Slosson a marqué 1 103 points d'affilé face au français Vignaux qui a alors marqué 1 531 points. Dix ans plus tard Schäfer marquait 3 000 points. De tels scores sont possibles car quand les joueurs parviennent à regrouper les billes près d'une bande et à les garder ensemble ils peuvent avoir de très longues séries. Cette technique s'appelle l'américaine. Les joueurs ne pourraient pas jouer pendant des heures sans manquer s'ils devaient faire face à de nombreux points difficiles.

Pour empêcher les joueurs d'utiliser l'américaine —et ainsi éviter les parties fastidieusement longues— de nouvelles règles ont été introduites. Dans les jeux de cadres il est interdit de laisser les billes dans un certain secteur de la table pendant plus d'un ou deux coups ; de longues séries (centaines de points) sont encore possible grâce au remplacement. Une autre manière de compliquer le jeu est d'imposer au joueur de toucher un certain nombre de bandes avant que sa bille n'entre en contact avec la troisième bille.

Même si les règles sont très similaires, ces différents modes de jeu requièrent différentes qualités et se jouent très différemment. Par conséquent, de nombreux joueurs jouent seulement au trois-bandes ou seulement aux jeux de séries. Peu de joueurs ont le même niveau dans tous les modes de jeu. A cause des différences de règles entre les différents modes de jeu, les scores varient beaucoup : les meilleurs joueurs de trois-bandes au monde marquent rarement vingt points d'affilée alors qu'il est possible de marquer des milliers de points à la libre. On doit donc comparer ce qui est comparable. Dans ce qui suit les applications numériques viennent du cadre

---

\* Email address: m-bouville@imre.a-star.edu.sg

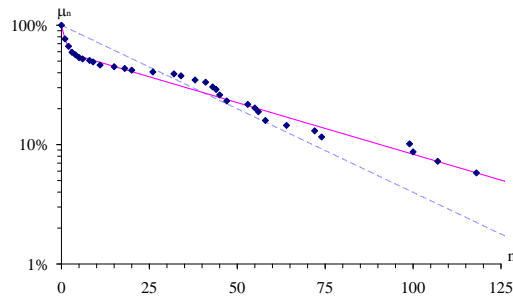


FIGURE 1 –  $\mu_n$ , la probabilité de marquer au moins  $n$  points, en fonction de  $n$  pour Villiers au cadre 47/2. Losanges : données ; ligne pointillée : modèle bernoullien ; ligne continue : modèle markovien.

47/2 (sauf dans la partie 6.2, qui est dédiée au trois-bandes). Le modèle peut naturellement être appliqué à tous les jeux de séries ; le seul critère est que les données soient commensurable.

## 2. Statistiques au billard

### 2.1. Le billard se plie aisément à un traitement statistique

Le billard a été étudié en détail d'un point de vue mécanique par Coriolis [3] puis plus récemment par Petit [4]. D'un autre côté, très peu a été fait dans le domaine des statistiques. C'est surprenant quand on considère à quel point le billard français se plie aisément à un traitement statistique. Contrairement au pool ou au snooker, le nombre de billes sur la table est constant. D'autre part, chaque coup rapporte le même nombre de points, alors qu'à la 8 ou à la 9 il n'y a qu'une bille qui ait vraiment de la valeur, au 14.1 tous les quinze coups on doit jouer un coup très différent, et au snooker il y a deux parties dans un match. Le billard carambole est essentiellement un régime permanent : quand on regarde les billes sur la table on ne peut pas dire si on est en train de jouer le troisième coup ou le centième.

### 2.2. La moyenne

A la fin du match une moyenne arithmétique peut être obtenue en divisant le nombre de points par le nombre de reprises — on peut aussi calculer des moyennes pour un tournoi, une année, etc. On peut ainsi établir des classements, même pour des joueurs qui n'ont jamais joué l'un contre l'autre (comme en athlétisme où on peut comparer temps et distances pour classer des athlètes qui ne se sont jamais affrontés). La moyenne est parfois utilisée pour établir des classements mais aussi (ce qui est peut-être plus important) elle a une grande importance psychologique pour les joueurs ; par exemple beaucoup de gens trouvent bizarre que le vainqueur d'un tournoi ne soit pas aussi le joueur qui a la moyenne la plus élevée.

### 2.3. Le billard carambole comme processus de Bernoulli

Malgré l'importance de la moyennes au billard, l'étude des statistiques du billard s'est limitée à supposer un processus de Bernoulli. Un tel modèle suppose que la probabilité de marquer sur un certain coup ne dépend pas des coups précédents. Soit  $\mu_n$  la probabilité de marquer au moins  $n$  points, avec  $n$  entier. Pour un processus de Bernoulli avec un taux de réussite  $\lambda < 1$ , pour tout  $n \geq 0$  le terme général de  $(\mu_n)$  est

$$\mu_n = \lambda^n. \quad (1)$$

La longueur moyenne d'une série —ci-après appelée « la moyenne »— vaut

$$m = \frac{\lambda}{1 - \lambda}. \quad (2)$$

La figure 1 représente des résultats de Bernard Villiers au cadre 47/2, basés sur 69 séries complètes (les séries qui ont été interrompues parce que Villiers a atteint 200 points et gagné le match ne sont pas comptabilisées). Le modèle bernoullien (pointillés) donne des résultats très

différents des données (losanges) : il surestime l'importance des courtes séries et sous-estime celle des longues séries ( $n$  grand).

#### 2.4. L'importance du remplacement

Un processus de Bernoulli ne tient pas compte du remplacement, qui est très important dans les jeux de séries. La probabilité de marquer un point dépend de sa difficulté, qui n'est pas nécessairement aléatoire. Face à un coup difficile le joueur cherche surtout à marquer (on suppose que le joueur cherche à marquer le point sans chercher à défendre, c'est-à-dire sans se soucier de ce qui arriverait s'il manquait). Sur les coups faciles en revanche marquer n'est pas un problème et le joueur peut essayer de jouer d'une manière telle que le point sera marqué et la position finale des billes sera favorable. Ceci s'appelle le remplacement. Après un coup facile, on peut donc obtenir un autre coup relativement facile. Un joueur peut ainsi avoir une série de coups faciles et de très longues séries. De telles corrélations sont complètement absentes du modèle bernoullien décrit ci-dessus ; il ne peut donc décrire correctement les jeux de séries. Si quelques limitations du modèle bernoullien ont été considérées [5, 6] l'effet du remplacement n'a jamais été étudié. Afin de tenir compte des corrélations entre coups successifs on va modéliser le billard carambole comme un processus de Markov.

### 3. Le billard carambole comme processus de Markov

#### 3.1. Un processus de Markov instinctif

Si on considère qu'il existe  $N_0$  positions possibles des billes sur la table, le résultat d'un coup peut être un succès avec les billes finissant dans une des  $N_0$  configurations possibles ou un échec. La probabilité de marquer et la position des billes pour le prochain coup dépendent seulement de la position des billes au début du coup, pas de leur position antérieure. La succession des coups est donc un processus de Markov à  $N_0 + 1$  états. Malheureusement, comme il existe une infinité de positions possibles, les probabilités des transitions ne peuvent être toutes obtenues. Concrètement, un tel modèle est inapplicable. Pour simplifier le modèle et le rendre utilisable, on peut regrouper des positions similaires pour n'avoir que  $N$  types de positions. Si  $N$  est assez petit, les paramètres sont suffisamment peu nombreux pour pouvoir être calculés et interprétés concrètement.

#### 3.2. Un processus de Markov utilisable

Soit  $\mathbf{P}_n$  le vecteur à  $N$  dimensions dont la composante  $i$  est la probabilité que les billes soient dans une position de type  $i$  après le coup  $n$  (on utilise « coup de type  $i$  » et « position [des billes sur la table] de type  $i$  » comme des synonymes). Pour tout  $n \geq 0$  la suite  $(\mathbf{P}_n)$  obéit à  $\mathbf{P}_{n+1} = \mathbf{K} \mathbf{P}_n$ . L'élément  $k_{ij}$  de  $\mathbf{K}$ , matrice  $N \times N$ , est la probabilité de marquer sur un coup de type  $j$  et que le prochain coup soit de type  $i$ .  $\mathbf{K}$  est *apparentée* à une matrice de transitions : l'état correspondant à un manqué n'est pas inclus car les composantes afférentes peuvent être obtenues simplement à partir de  $\mathbf{K}$  (la probabilité de manquer un coup de type  $i$  est  $1 - \kappa_i$ , où  $\kappa_i = \sum_j k_{ji}$  est la probabilité de marquer sur un coup de type  $i$ ). Pour tout  $n \geq 0$ ,  $\mathbf{P}_n = \mathbf{K}^n \mathbf{P}_0$ .

Dans ce qui suit on utilise  $N = 2$ . Soit  $p_0$  la probabilité que le premier point joué par le joueur soit de type 1. Soient  $\rho_1 \leq \rho_2$  les valeurs propres de  $\mathbf{K}$  et  $\mathbf{V}_1$  et  $\mathbf{V}_2$  les vecteurs propres associés. Pour tout  $n \geq 0$ ,  $\mathbf{P}_n$  s'écrit

$$\mathbf{P}_n = \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{pmatrix} \begin{pmatrix} (\rho_1)^n & 0 \\ 0 & (\rho_2)^n \end{pmatrix} \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{pmatrix}^{-1} \begin{pmatrix} p_0 \\ 1 - p_0 \end{pmatrix}. \quad (3)$$

#### 3.3. La probabilité de marquer au moins $n$ points

La somme des composantes de  $\mathbf{P}_n$  est la probabilité d'avoir les billes dans n'importe quelle position après le coup  $n$ , c'est-à-dire la probabilité d'avoir marqué au moins  $n$  points. La probabilité de marquer au moins  $n$  points,  $\mu_n$ , est donc la somme des composantes de  $\mathbf{P}_n$ . Pour tout  $n \geq 0$ ,  $\mu_n$  est de la forme

$$\mu_n = (1 - Y)(\rho_1)^n + Y(\rho_2)^n, \quad (4)$$

$Y$  est un scalaire connu en fonction de  $\mathbf{V}_1$ ,  $\mathbf{V}_2$  et  $p_0$ . La figure 1 représente  $\mu_n$  sur une échelle semi-logarithmique en fonction de  $n$  pour Bernard Villiers au cadre 47/2. La ligne continue de la figure 1 est clairement en bien meilleur accord avec les données que la ligne pointillée : le processus de Markov est une bien meilleure description des jeux de cadres que le processus de Bernoulli.

La moyenne,  $m$ , est la somme sur  $n$  de  $\mu_n$ . En utilisant que  $\sum_i k_{ij} < 1$ , la norme de  $\mathbf{KV}$  est  $\|\mathbf{KV}\| = \sum_i |\sum_j k_{ij} v_j| \leq \sum_{ij} k_{ij} |v_j| < \sum_j |v_j| = \|V\|$ . Par définition de  $V_i$  et de  $\rho_i$ , on a  $\|\mathbf{KV}_i\| = |\rho_i| \|V_i\|$ . Et comme  $\|\mathbf{KV}_i\| < \|V_i\|$ , on trouve que  $|\rho_i| < 1$ . Donc la somme sur  $n$  de  $(\rho_i)^n$  converge et

$$m = (1 - Y) \frac{\rho_1}{1 - \rho_1} + Y \frac{\rho_2}{1 - \rho_2}. \quad (5)$$

### 3.4. Les différents types de positions

On a considéré qu'il existe deux types de coups, défini la matrice  $\mathbf{K}$  et utilisé ses vecteurs propres  $\rho_1$  et  $\rho_2$  pour calculer  $\mu_n$  dans l'équation (4). Mais on n'a jamais vraiment spécifié ce que sont ces deux types de coups. Ils ne peuvent pas être aléatoires, sinon le processus de Markov se réduirait à un processus de Bernoulli. Les types de coups doivent avoir deux propriétés. (1) Comme le modèle se base sur les probabilités de marquer, les types de coups devraient être définis en fonction de ces probabilités. (2) Les positions de type  $i$  doivent être telles qu'elles sont généralement suivies d'une autre position de type  $i$ .

Les positions de type 2 correspondent aux positions « faciles » et les positions de type 1 aux positions « difficiles ». Par exemple, à la libre les coups de type 2 seraient des positions d'américaine. Cette définition satisfait aux deux conditions : (1) les types de coups sont définis en termes de probabilités de marquer (type 1 : probabilité basse, type 2 : probabilité élevée) et (2) les coups faciles autorisent le joueur à se concentrer sur le remplacement, si bien qu'ils sont généralement suivis d'un autre coup facile.

## 4. Similarités et différences entre joueurs

On aimerait utiliser le modèle pour comparer des joueurs de niveaux différents : les meilleurs sont-ils meilleurs en tout ou bien y a-t-il *une* différence qui les rend meilleurs ? Par exemple, est-ce que les meilleurs ont des scores plus élevés parce qu'ils ne font jamais de séries très courtes ? Ou bien est-ce qu'ils font plus souvent des séries très longues ? Une telle comparaison est difficile parce que la principale différence entre meilleurs et moins bons joueurs est la différence de leur moyenne : il faut parvenir à annuler cette différence pour que les effets d'« ordre supérieur » soient visibles.

### 4.1. Dédimensionnalisation

Dans le modèle bernoullien —équation (1)—,  $\mu_n = \exp(n \ln \lambda)$ . En posant  $\nu = (-\ln \lambda)n$ ,  $\mu_n$  est égal à  $\exp(-\nu)$ . Si  $m$  est grand,  $\nu \sim n/m$  : la longueur des séries est ainsi mesurée en pourcentage de la moyenne. On peut ainsi comparer des joueurs ayant moyennes différentes : si on trace  $\mu_n$  en fonction de  $(-\ln \lambda)n$  pour plusieurs joueurs, tous les résultats se trouveront sur la même courbe.

La probabilité de marquer  $\lambda$  vaut  $m/(m+1)$  dans le modèle bernoullien [voir équation (2)]. Dans le modèle markovien il n'y a pas de probabilité de marquer unique. On définit le paramètre  $\lambda$  par  $m/(m+1)$  ;  $\lambda$  est la probabilité de marquer dans un modèle bernoullien qui donnerait une moyenne de  $m$ . On trace  $\mu_n$  en fonction de  $(-\ln \lambda)n$  [figure 2(a)]. Les résultats sont similaires même pour des joueurs de niveaux très différents (les données sont pour des joueurs de moyenne comprise entre 5 et 35).

Cependant, la figure 2(a) présente des variations, en particulier pour les grands  $n$ . Afin de déterminer si ceci est corrélé au niveau de jeu des joueurs ou bien provient de différences entre joueurs de même niveau, on regroupe les 25 joueurs de la figure 2(a) en trois joueurs composites. On utilise les scores de joueurs de moyenne supérieure à 25 pour créer un « joueur » ayant joué environ 400 reprises pour une moyenne de 30,4. De même on utilise les scores de joueurs ayant

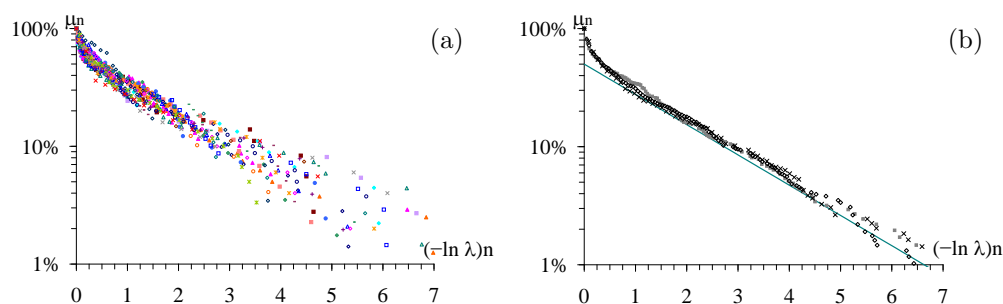


FIGURE 2 – La probabilité de marquer au moins  $n$  points,  $\mu_n$ , en fonction de  $(-\ln \lambda)n$  où  $\lambda = m/(m+1)$ . (a) : données de 25 joueurs français de cadre 47/2 de moyenne comprise entre 5 et 35. (b) : trois joueurs composites (symboles pleins : joueurs de moyenne supérieure à 25, symboles vides :  $15 < m < 25$ , croix :  $m < 15$ ). La droite en (b) est  $-0,59(-\ln \lambda)n + \ln 0,50$ .

entre 15 et 25 de moyenne (700 reprises,  $m = 20,5$ ) et de joueurs de moyenne inférieure à 15 (500 reprises,  $m = 10,7$ ). La figure 2(b) représente les scores de ces trois « joueurs ». Il n’y a pas de différence notable entre eux. Les variations de la figure 2(a) proviennent donc de variations entre joueurs de même niveau et non de différences entre meilleurs et moins bons.

#### 4.2. $\rho_2$ en fonction de la moyenne

Dans les cas qui nous intéressent,  $\rho_1$  est nettement inférieur à  $\rho_2$  (sinon le processus serait quasi bernoullien et un modèle markovien ne serait pas nécessaire). Donc, pour les grands  $n$ ,  $(\rho_1)^n = o(\rho_2)^n$  et  $\ln \mu_n \sim \ln Y + n \ln \rho_2$ . La figure 2(b) montre qu’effectivement  $(\ln \mu_n)$  a une asymptote pour les grandes valeurs de  $(-\ln \lambda)n$ . Soit  $a$  la valeur absolue de sa pente,  $a = (\ln \rho_2)/(\ln \lambda)$ . Ainsi  $\rho_2$  peut être obtenu en fonction de la moyenne :

$$\rho_2 \approx \lambda^a = \left( \frac{m}{m+1} \right)^a \quad (6)$$

pour ces trois « joueurs ». La figure 3(b) représente  $\rho_2$  en fonction de  $\lambda$  : la droite, qui correspond à  $\lambda^{0.59}$ , confirme que l’équation (6) avec  $a = 0.59$  est une bonne approximation de  $\rho_2$ . Comme la plupart des points sont marqués sur des coups faciles, il n’est pas surprenant qu’il existe une corrélation entre  $\rho_2$  et la moyenne (donc  $\lambda$ ). De la figure 2(b) on obtient aussi que  $Y \approx 0.50$ . Les valeurs de  $a$  et  $Y$  ne dépendent pas du niveau des joueurs mais elles dépendent probablement du mode de jeu ; les valeurs  $a = 0.59$  et  $Y = 0.50$  pourraient ainsi être une sorte de « signature » du cadre 47/2. (D’autres travaux seront nécessaires pour trouver les valeurs de  $a$  et  $Y$  pour les autres modes de jeu.)

#### 4.3. $\rho_1$ en fonction de la moyenne

On peut écrire  $\rho_1$  en fonction de  $\rho_2$ ,  $Y$  et  $m$  en utilisant l’équation (5). Comme  $m = \lambda/(1-\lambda)$  et  $\rho_2 = \lambda^a$  la valeur propre  $\rho_1$  vaut

$$\rho_1 = 1 - \frac{(1-Y)(1-\lambda)}{1 - Y \frac{1-\lambda}{1-\lambda^a}}. \quad (7)$$

Les paramètres  $a$  et  $Y$  sont connus à partir de la figure 2(b). La figure 3(a) montre que l’équation (7) est une médiocre approximation de  $\rho_1$ . Comme les coups difficiles contribuent peu à la moyenne, il n’est pas étonnant que la corrélation entre  $\rho_1$  et  $\lambda$  soit faible.

#### 4.4. Différences entre meilleurs et moins bons joueurs

Si l’adversaire laissait toujours un coup facile la moyenne serait d’environ  $\rho_2/(1-\rho_2)$ . Soit  $m_2$  la contribution à la moyenne des séries commençant par un coup de type 2. En utilisant les équations (5) et (6), on trouve que la contribution relative des séries commençant par un coup

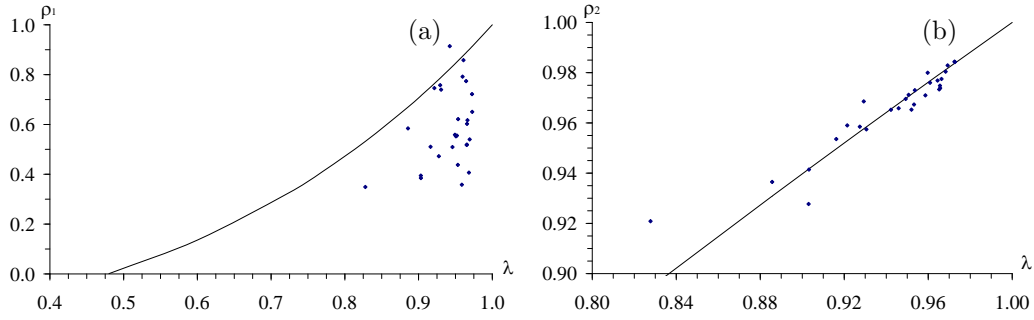


FIGURE 3 – (a)  $\rho_1$  et (b)  $\rho_2$  en fonction de  $\lambda$ . La courbe en (a) correspond à l'équation (7) et la droite en (b) à l'équation (6) avec  $a = 0,59$  et  $Y = 0,50$ .

facile est

$$\frac{m_2}{m} \approx (1 - p_0) \frac{\lambda^a (1 - \lambda)}{\lambda (1 - \lambda^a)}. \quad (8)$$

On peut facilement montrer que  $m_2/m$  est une fonction décroissante de  $\lambda$ . Les joueurs les moins bons ( $\lambda$  petit) marquent donc principalement quand leur adversaire leur laisse un coup facile alors que les meilleurs marquent aussi sur des séries faites à partir d'une position difficile, c'est-à-dire qu'ils sont moins sensibles à la stratégie de leur adversaire.

Pour résumer les différences entre meilleurs et moins bons joueurs : (i)  $\rho_2$  est une indication du niveau du joueur ( $\rho_2$  est corrélée à la moyenne), (ii) il n'y a guère de corrélation entre  $\rho_1$  et la moyenne du joueur :  $\rho_1$  dépend peut-être des particularités du joueur (forces, faiblesses, tactique) et (iii) les meilleurs joueurs marquent une plus grande proportion de leurs points sur des séries débutant par un coup difficile.

## 5. Calcul des probabilités de marquer

On aimerait calculer les probabilités de marquer d'un joueur (sa matrice  $K$ ) à partir de ses scores.

### 5.1. Un système sous-défini

Malheureusement, le système est sous-défini : les cinq inconnues (les quatre éléments de  $K$  et  $p_0$ ) ne débouchent que sur trois paramètres mesurables ( $m$  et les valeurs propres  $\rho_1$  et  $\rho_2$ ). Grâce à l'équation (5),  $Y$  peut être obtenu en fonction de  $\rho_1$ ,  $\rho_2$  et  $m$ . On peut ainsi récrire l'équation (4) pour exprimer  $(\mu_n)$  en fonction de  $\rho_1$ ,  $\rho_2$  et  $m$ . La moyenne  $m$  est connue grâce aux feuilles de matches et on trouve les valeurs propres  $\rho_1$  et  $\rho_2$  en minimisant

$$\sigma^2 = \sum_n \frac{\left( \mu_n^{\text{Eq. (4)}} - \mu_n^{\text{data}} \right)^2}{\mu_n^{\text{data}}}. \quad (9)$$

(Minimiser des fonctions similaires, par exemple en ne divisant pas par  $\mu_n^{\text{data}}$  ou en divisant par son carré, donne des résultats similaires.)

Comme le système est sous-défini, on ne peut pas déterminer  $K$  uniquement. Tout ce qu'on peut faire est écrire  $K$  en fonction de  $\rho_1$ ,  $\rho_2$ ,  $m$  et de deux paramètres libres. Ces deux paramètres ne peuvent pas prendre n'importe quelle valeur : pour tout couple  $(i, j)$  on doit avoir  $0 < k_{ij} < 1$  et  $k_{1j} + k_{2j} < 1$ . Ces contraintes autorisent tout de même de larges intervalles de valeurs possibles pour les paramètres libres et donc pour  $K$ . Par exemple pour Bernard Villiers au cadre 47/2 on peut avoir

$$K = \begin{pmatrix} 40,7 \% & 0,0 \% \\ 0,0 \% & 98,0 \% \end{pmatrix} \quad \text{ou} \quad K = \begin{pmatrix} 43,5 \% & 4,8 \% \\ 32,2 \% & 95,2 \% \end{pmatrix}. \quad (10)$$

Les incertitudes dans l'expression (10) proviennent principalement du large intervalle de valeurs possibles pour  $p_0$  : si on fixait  $p_0$ ,  $K$  serait connue plus précisément.

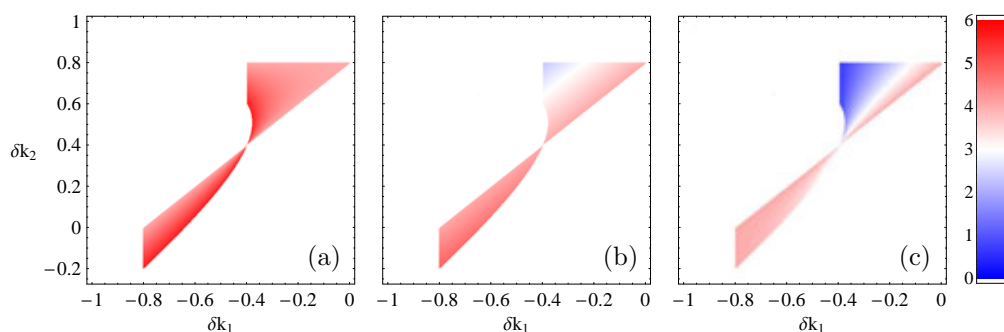


FIGURE 4 – (couleur) La moyenne pour  $\rho_1 = 0,4$  et  $\rho_2 = 0,8$ . En (a) la proportion de coups difficiles laissés par l’adversaire est de  $p_0 = 0$ , en (b)  $p_0 = 1/2$  et en (c)  $p_0 = 1$ .

## 5.2. La valeur de $p_0$

Il existe un spectre continu de difficultés et la frontière entre coup facile et coup difficile est dans une certaine mesure arbitraire. La valeur de  $p_0$  est liée à cette frontière : fixer  $p_0$  détermine quel niveau de facilité est requis pour qu’une position soit de type 2. Si  $p_0$  est petit la plupart des coups sont faciles alors que la plupart des coups sont difficiles si  $p_0$  est grand. Dans ce qui suit, on appellera « coups faciles » les coups plus faciles que la médiane de la probabilité de marquer. Ainsi, par définition 50 % des coups sont faciles et 50 % difficiles, c’est-à-dire  $p_0 = 1/2$ . Pour Bernard Villiers jouant au cadre 47/2,

$$\mathbf{K} = \begin{pmatrix} 40,9 \% \pm 0,2 \% & 1,2 \% \pm 1,2 \% \\ 11,5 \% \pm 1,2 \% & 97,8 \% \pm 0,2 \% \end{pmatrix}. \quad (11)$$

Des incertitudes demeurent vu qu’après avoir fixé  $p_0$  il y a encore un paramètre libre. Ces incertitudes sont tout de même bien plus faibles que celles de l’expression (10), surtout pour  $k_{21}$ .

La probabilité de marquer sur les coups faciles,  $k_{12} + k_{22}$ , est proche de un et peut même être égale à un : comme on pouvait s’y attendre il est quasi impossible que Villiers manque sur un coup facile. De plus la plupart des coups faciles débouchent sur un autre coup facile ( $k_{22} \geq 97.6$  %). Comme prévu le taux de réussite sur les coups difficiles est bien plus bas :  $k_{11} + k_{21} < 54$  %. Quand Villiers marque à partir d’une position difficile, il obtient une autre position difficile environ 80 % du temps.

## 5.3. Les joueurs peuvent calculer leur probabilités de marquer en ligne

La page <http://billiards.mathieu.bouville.name/biMar/> permet aux joueurs de calculer leur matrice  $\mathbf{K}$  à partir de leurs scores. Le programme minimise l’équation (9) par la méthode du gradient conjugué pour obtenir les valeurs propres. Il en déduit l’intervalle possible des éléments de la matrice  $\mathbf{K}$  comme on a fait pour Villiers dans l’expression (11). L’utilisateur fournit des données en deux colonnes, la première est un nombre de points par reprise et la seconde est le nombre de fois où ce score a été obtenu. Ces données ainsi que le résultat, la probabilité d’obtenir un coup facile/difficile après un coup facile/difficile, sont assez simples. Le joueur n’a pas besoin de connaître le modèle ni les mathématiques pour obtenir ce résultat.

## 6. Applications et cas particuliers

### 6.1. Stratégie de l’adversaire

Le premier coup joué par un joueur est le coup laissé par son adversaire. Il n’est pas forcément aléatoire : si l’adversaire joue défensivement, il laissera souvent les billes dans une position difficile. La figure 4 représente la moyenne si  $\rho_1 = 0,4$  et  $\rho_2 = 0,8$ , pour trois types d’adversaires. Comme la moyenne n’est pas fixée il y a deux paramètres libres ; on utilise  $\delta k_1 = k_{21} - k_{11}$  et  $\delta k_2 = k_{22} - k_{12}$ . Pour la figure 4(a) le premier coup que le joueur doit jouer est toujours facile. Dans le cas de la figure 4(c) l’adversaire laisse seulement des coups difficiles et la figure 4(b) correspond à un joueur « médian » qui laisse 50 % de coups faciles et 50 % de coups difficiles.

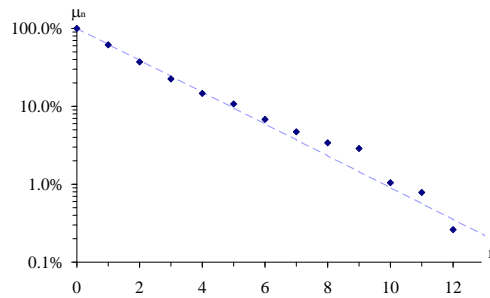


FIGURE 5 –  $\mu_n$ , la probabilité de marquer au moins  $n$  points, en fonction de  $n$  pour Raymond Ceulemans aux championnats du monde de trois-bandes 1978. Losanges : données ; ligne pointillée : modèle bernoullien.

Si  $\delta k_1 = -0,4$  et  $\delta k_2 = 0,8$  la moyenne dépend fortement de la stratégie de l’adversaire. Quand  $\delta k_1 = -0,8$  et  $\delta k_2 = 0$  en revanche la moyenne est indépendante de ce que fait l’adversaire. Ceci fournit des informations supplémentaires : si pour un joueur donné on connaît  $\rho_1$  et  $\rho_2$  et comment il se comporte contre divers joueurs il est possible de réduire le nombre de valeurs possibles de  $\mathbf{K}$ . Si par exemple on sait qu’un joueur avec  $\rho_1 = 0,4$  et  $\rho_2 = 0,8$  est très sensible à la stratégie de son adversaire on déduira que  $\delta k_1 \approx -0,4$  et  $\delta k_2 \approx 0,8$ .

## 6.2. Cas du trois-bandes

La figure 5 représente  $\ln \mu_n$  en fonction de  $n$  pour Raymond Ceulemans au championnat du monde de trois-bandes 1978 où il a joué 382 reprises, remportant tous ses matches et le titre. Raymond Ceulemans est probablement le meilleur joueur de tous les temps, il a plus de 100 titres (mondiaux, européens et belges) et a remporté 17 des 18 championnats du monde de trois-bandes organisés entre 1963 et 1980. Même pour Ceulemans au sommet de sa carrière le trois-bandes est difficile et, à la différence de ce qui a été observé sur la figure 1 au cadre 47/2, les résultats de Ceulemans sont proches d’un processus de Bernoulli, c’est-à-dire des résultats d’un joueur qui ne jouerait pas du tout le remplacement. En se basant sur près de 400 reprises on ne peut donc pas montrer qu’il existe une corrélation entre coups successifs et donc que Ceulemans joue le remplacement (alors que quelques dizaines de reprises suffisent aux jeux de séries). Le modèle bernoullien semble donc suffire dans le cas du trois-bandes.

## 6.3. Un processus de Markov avec $N$ types de coups

L’équation (4) montre que, si  $N = 2$ ,  $(\mu_n)$  est la somme de deux suites géométriques. Dans le cas général,  $(\mu_n)$  est la somme de  $N$  suites géométriques. Il y a  $N^2 + N - 1$  inconnues ( $\mathbf{K}$  a  $N^2$  éléments et  $\mathbf{P}_0$  en a  $N - 1$ ) dont  $N + 1$  peuvent être obtenues à partir des feuilles de matches ( $N$  valeurs propres et la moyenne). Ceci laisse  $N^2 - 2$  paramètres libres : deux paramètres libres si  $N = 2$  mais sept si  $N = 3$ . Sur la figure 1 il ne semble pas y avoir plus de deux exponentielles : on n’a pas besoin de choisir une valeur de  $N$  supérieure à 2. C’est heureux car sinon on aurait dû se coltiner sept paramètres libres (ou plus), ce qui n’est pas concrètement possible.

## 7. Conclusion

On a présenté un nouveau modèle du billard carambole qui tient compte du remplacement. En modélisant les jeux de séries comme un processus de Markov la probabilité de marquer un point est corrélée à la difficulté du coup précédent. La probabilité de marquer au moins  $n$  points est alors une somme de deux suites géométriques (de raisons  $\rho_1 < \rho_2$ ). Le processus de Bernoulli généralement utilisé jusqu’à maintenant est un cas particulier du présent modèle avec une seule suite géométrique. Le modèle markovien est une bien meilleure description des jeux de cadres que le modèle bernoullien. En utilisant ce modèle markovien on a pu établir des différences entre joueurs de différents niveaux. La valeur de la plus grande raison,  $\rho_2$ , est plus élevée pour les meilleurs joueurs. D’un autre côté il n’y a guère de corrélation entre la plus petite raison,  $\rho_1$ , et la moyenne du joueur. Enfin, les meilleurs joueurs marquent une



plus grande proportion de leurs points sur des séries commençant par une position difficile. On a aussi montré comment obtenir les probabilités de marquer des joueurs à partir de leurs feuilles de matches. Il existe une certaine incertitude sur les résultats parce que le système est sous-défini, de telle sorte qu'on obtient des intervalles plutôt que des valeurs exactes. La page <http://billiards.mathieu.bouville.name/biMar/> permet aux joueurs d'utiliser ce modèle sans difficulté en utilisant leurs propres scores.

### Remerciements

Je tiens à remercier A. Allenic pour ses conseils avisés ainsi que J.-L. Frantz, J.-M. Fray et R. Jewett pour m'avoir fourni des feuilles de matches.

### Glossaire

**américaine** : à la libre, le joueur rassemble les billes à proximité d'une bande afin de n'avoir que des coups faciles à jouer et d'obtenir de très longues séries.

**bandes (une bande, trois bandes)** : modes de jeu où la bille du joueur doit entrer en contact avec un certain nombre de bandes (une ou trois) avant de conclure le point.

**cadres** : trois modes de jeu (47/2, 47/1 et 71/2) où les billes ne peuvent pas être laissées dans la même partie du billard pendant plus d'un ou deux coups.

**défense** : au lieu d'essayer de marquer à tout prix, le joueur s'assure qu'il ne laissera pas une position facile à son adversaire s'il manque (même si ça signifie une plus grande probabilité de manquer).

**feuille de match** : document sur lequel sont enregistrés les scores pendant un match, reprise par reprise.

**moyenne** : nombre moyen de points marqués par reprise. Utilisée officiellement et/ou officieusement pour classer les joueurs.

**partie libre (ou libre)** : mode de jeu le plus simple dans lequel un point est marqué si la bille du joueur touche les deux autres billes, sans autre contrainte.

**remplacement** : fait d'essayer non seulement de marquer le point mais aussi d'obtenir une position favorable pour le prochain coup afin de n'avoir que des coups faciles à jouer et des scores élevés. L'américaine en est un exemple.

**série** : coups successifs sans manquer. Nombre de points ainsi marqués, *une série de 10*.

### Bibliographie

- [1] BOUVILLE, M. (2007). Position play in carom billiards as a Markov process, *Journal of Quantitative Analysis in Sports* **3**(4), art. 4. Cet article est disponible en ligne à cette adresse : <http://www.bepress.com/jqas/vol3/iss4/4/>.
- [2] BOUVILLE, M. (2005). Le français, le suisse, le russe et l'américaine. Disponible en ligne à <http://billiards.mathieu.bouville.name/suisse-russe-americaaine.pdf>.
- [3] CORIOLIS, G. G. (1835). *Théorie mathématique des effets du jeu de billard*. Carilian-Gœury, Paris, France (réédité en fac-similé par Phénix Éditions en 2003).
- [4] PETIT, R. (1996). *Billard, théorie du jeu*. Chiron (Paris, France).
- [5] FRAY, J.-M. (2002). A propos de la moyenne au billard carambole. In F. Caudron (ed.) : *Le billard en expansion*. Deleye (Ledegem, Belgium).
- [6] BOUVILLE, M. (2004). Statistiques des jeux de séries au billard carambole. Disponible en ligne à <http://billiards.mathieu.bouville.name/jeux-de-serie.pdf>.