Statistiques des jeux de séries au billard carambole

Mathieu Bouville

m-bouville @ imre.a-star.edu.sg

Résumé. A la libre ou aux jeux de cadre, certains matches peuvent ne durer que quelques reprises. Quel est l'effet sur le résultat du match ? C'est-à-dire : la probabilité que le meilleur gagne est-elle la même sur des matches en une reprise et des matches en une dizaine de reprises ? D'autre part la moyenne calculée sur des matches relativement courts est-elle une bonne approximation de la moyenne réelle ?
Cet article montre que pour que la moyenne ait un sens, il faut que la partie soit suffisamment longue. Elle doit donc se jouer en au moins trois ou quatre reprises. Sinon la moyenne du vainqueur est sous-estimée. Quand on calcule la moyenne comme le nombre de points réussis divisé par le nombre de points manqués (au lieu de diviser par le nombre de reprises), le résultat est meilleur mais pas parfait.

1. Introduction

Le championnat de Belgique 2005 à la partie libre se jouait en matches de 400 points. Sur les dix matches, quatre ont été joués en une seule reprise, y compris deux matches nuls 400-400 en une reprise. Patrick Niessen a fait deux matches nuls en marquant 400 points en une reprise. Dans son match contre Robby Sonk, Johan Claessen a marqué les 400 points en douze reprises et en a retiré plus de points que Niessen pour ses 400 points sur mouche. De telles situations peuvent toujours se produire, mais elles sont plus fréquentes dans le cas de parties en peu de reprises. On peut se demander si le résultat final du championnat aurait été différent avec des parties plus longues.

Cet article traite des matches en une ou deux reprises. Son but est d'établir si la distance a une influence sur l'issue du match. Si c'est le cas, combien de reprises faut-il au minimum ? Il est d'autre part généralement admis (sauf par les règlements) que la moyenne du vainqueur est sous-estimée aux jeux de série. La moyenne calculée comme le nombre de points réussis divisé par le nombre de points *manqués* est-elle plus fiable que la moyenne habituelle ?

Dans tout ce qui suit, on considère des matches avec reprise égalisatrice. On dira qu'un joueur vaut 100 de moyenne ou bien qu'il a une MG de 100 pour décrire le niveau réel moyen d'un joueur. Il n'est donc pas incohérent qu'un joueur ait une moyenne sur le match très différente de sa MG ou qu'un joueur de 100 de MG batte un joueur à 200 de MG.

La partie 2 est le développement mathématique ; sa lecture n'est pas indispensable pour comprendre la partie 3 qui donne les résultats.

2. Modélisation du match

2.1. *Principales définitions*

On appelle d la distance du match.

La probabilité de marquer au moins i points est $\mu_i = p^i$ où p est la "probabilié moyenne" de réussir un point †. C'est la probabilité de marquer moyennée sur tous les coups possibles. Certains coups sont faciles, d'autres pas, si on fait la moyenne on obtient la "probabilité moyenne". Si on joue une partie en une infinité de reprises, le nombre moyen de points par reprise va converger vers une valeur qu'on appellera m. m est la moyenne, elle vaut

$$m = \frac{p}{1-p}. \tag{1}$$

On définit $F_n(i)$ comme la probabilité d'avoir exactement i points ($i < d$) à la fin de la reprise n.

†Bouville (2005a) montre que $\mu_i = p^i$ n'est valable que sous certaines hypothèses (voir aussi Bouville (2005b)). Cependant l'utilisation d'une version améliorée de (μ_i) rend les calculs trop compliqués pour qu'on puisse l'utiliser ici.

On définit G_n comme la probabilité de gagner en n reprises. C'est donc la probabilité d'avoir d points à la fin de la reprise n mais moins de d points à la fin de la reprise $n-1$ (sinon victoire en $n-1$ reprises).

2.2. Probabilité de victoire

Un joueur qui marque i points en n reprises marque i fois et manque n fois, il joue donc $i+n$ fois. Si la probabilité de marquer est p alors la probabilité de marquer i fois et de manquer n fois est $p^i(1-p)^n$. Le nombre de manières d'obtenir ce résultat est le nombre de permutation des $n-1$ manqué parmi $i+n-1$ coups : C_{n+i-1}^{n-1} (le dermier manqué est forcément le dernier point). On a donc, pour tout $n \geq 1$,

$$F_n(i) = (1-p)^n p^i C_{i+n-1}^{n-1} = (1-p)^n p^i \frac{(i+1)(i+2)\ldots(i+n-1)}{(n-1)!}.$$

G_n est similaire à F_n : pour $n \geq 2$,

$$G_n = \sum_{i=0}^{d-1} F_{n-1}(i) p^{d-i}. \tag{2}$$

D'où

$$G_n = (1-p)^{n-1} p^d \sum_{i=0}^{d-1} f_{n-1}(i) = (1-p)^{n-1} p^d f_n(d-1). \tag{3}$$

Dans la pratique on pourra calculer G_n par récurrence :

$$G_n = (1-p)\frac{f_n(d-1)}{f_{n-1}(d-1)} G_{n-1} = (1-p)\frac{d+n-2}{n-1} G_{n-1}.$$

2.3. Dédimensionnalisation

Intuitivement on se dit qu'un joueur de moyenne m qui joue une partie en d points devrait finir en environ d/m reprises, que ce soit un joueur à 25 de moyenne qui fait une partie en 100 points ou un joueur à 100 de moyenne qui fait une partie en 400 points. Or ce ratio n'apparait nulle part dans les calculs précédents. On va donc ici essayer de réécrire G_n en fonction de d/m.

On appelle r le ratio d/m. En utilisant l'équation (1), la probabilité de marquer, p, peut être écrite en fonction de la moyenne, $p = m/(m+1)$. D'après l'équation (3),

$$G_n = \frac{1}{(n-1)!}\left(\frac{m}{m+1}\right)^{mr} \frac{d(d+1)\ldots(d+n-2)}{(m+1)^{n-1}}. \tag{4}$$

2.3.1. *Approximation de (G_n) à l'ordre 0*

En multipliant par r^{n-1} dans l'équation (4), on obtient

$$G_n = \frac{1}{(n-1)!}\left(\frac{m}{m+1}\right)^{mr} r^{n-1} \frac{\left(1+\frac{1}{d}\right)\ldots\left(1+\frac{n-2}{d}\right)}{\left(1+\frac{1}{m}\right)^{n-1}}.$$

En utilisant que

$$\left(\frac{m}{m+1}\right)^{mr} = e^{-r}\left[1+\frac{r}{2m}+o\left(\frac{1}{m}\right)\right],$$

on a, à l'ordre 0 en $1/d$,

$$G_n \approx e^{-r}\frac{r^{n-1}}{(n-1)!}. \tag{5}$$

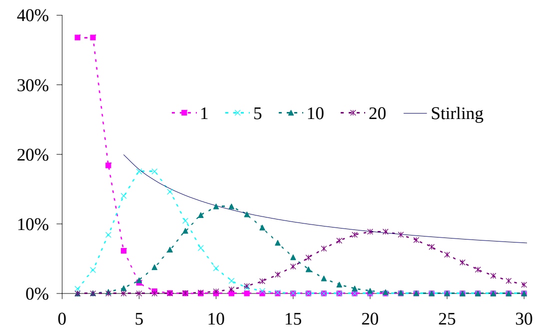

Figure 1. G_n (approximée) en fonction de n pour diverses valeurs de r. Aussi l'approximation du maximum en utilisant Stirling.

Ceci ne dépend que de n et r (ni de m ni de d). Donc dans les cas (jeux de série) où la moyenne du joueur et la distance sont grandes, G_n ne dépend que du ratio distance sur moyenne.

2.3.2. Variations de l'approximation de la suite (G_n)

L'équation (5) permet d'écrire

$$G_{n+1} > G_n \Leftrightarrow \frac{e^{-r}r^{n-1}}{(n-1)!}\left(\frac{r}{n}-1\right) > 0.$$

(G_n) a un maximum en $n = E(r)+1$ (si r est entier $G_r = G_{r+1}$.) La durée la plus probable du match est donc d'environ r ou $r+1$ reprises. En utilisant l'approximation de Stirling, ce maximum vaut

$$G_{max} \approx \frac{1}{\sqrt{2\pi r}}.$$

La figure 1 représente G_n (d'après l'équation (5)) en fonction de n. La courbe continue sur la figure 1 est le niveau du maximum dans l'approximation de Stirling.

2.3.3. Somme des G_n

G_n est la probabilité de marquer les d points en n reprises. La probabilité de finir la partie en *au plus* n reprises est $\sum_{i=1}^{n} G_i$. En utilisant l'approximation de G_n (équation (5)), on a

$$\sum_{i=1}^{n} G_i \approx e^{-r}\sum_{i=0}^{n-1}\frac{r^i}{i!},$$

où l'on reconnait de développement de l'exponentielle de r. En utilisant un développement de Taylor avec reste intégral on obtient

$$\sum_{i=0}^{n-1}\frac{r^i}{i!} = e^r - e^r\int_0^r \frac{t^{n-1}e^{-t}}{(n-1)!}\,dt,$$

et finalement

$$1-\sum_{i=1}^{n} G_i \approx \int_0^r \frac{t^{n-1}e^{-t}}{(n-1)!}\,dt \approx \frac{r^n}{(n-1)!}\int_0^1 t^{n-1}e^{-tr}\,dt. \quad (6)$$

Quand $n \to \infty$, $\sum_{i=1}^{n} G_i \to 1$, la probabilité de finir un jour vaut 1.

2.3.4. Nombre de reprises

Le nombre moyen de reprises est

$$\overline{rep} = \sum_{n=1}^{+\infty} \left(1 - \sum_{k=1}^{n-1} G_k \right)$$

pour un joueur beaucoup plus fort que son adversaire (s'il ne perd aucun match, le nombre de reprises ne dépend que de lui, pas de son adversaire). En utilisant l'approximation de G_n de l'équation (5), le nombre moyen de reprises est d'environ

$$\overline{rep} \approx 1 + \sum_{n=1}^{+\infty} \int_0^r \frac{t^{n-1} e^{-t}}{(n-1)!}\, dt \approx 1 + \int_0^r e^{-t} \left[\sum_{n=1}^{+\infty} \frac{t^{n-1}}{(n-1)!} \right] dt.$$

On obtient donc que

$$\overline{rep} \approx r + 1. \tag{7}$$

L'équation (7) donne une borne supérieure. En effet si le joueur joue contre un adversaire de son niveau le nombre de reprises sera en moyenne inférieur à $r+1$ car l'équation (7) inclue la possibilité d'une mauvaise partie (partie en un nombre élevé de reprise), dans un tel cas c'est vraisemblablement l'adversaire qui gagnerait. La partie ne se joue donc en beaucoup de reprises que si les deux joueurs jouent mal, ce qui a moins de chances d'arriver que le cas où un seul joueur joue mal. L'équation (7) surestime la probabilité d'un match long et surestime donc $\overline{rep}$.

2.3.5. Approximation de (G_n) au premier ordre

L'équation (5) donne (G_n) à l'ordre 0, à l'ordre 1 on a

$$G_n = \frac{e^{-r} r^{n-1}}{(n-1)!} \left[1 + \frac{(n-r-3/2)^2 - r - 1/4}{2d} \right] + o\left(\frac{1}{d}\right). \tag{8}$$

La somme des termes de (G_n) est au premier ordre

$$\sum_{i=1}^{n} G_i = e^{-r} \sum_{i=0}^{n-1} \frac{r^i}{i!} - e^{-r} \frac{r^n}{n!} \frac{n(n-r-1)}{2d} + o\left(\frac{1}{d}\right),$$

d'où

$$1 - \sum_{i=1}^{n} G_i = \frac{e^{-r} r^n}{(n-1)!} \left[\int_0^1 t^{n-1} e^{(1-t)r}\, dt + \frac{n-r-1}{2d} \right] + o\left(\frac{1}{d}\right).$$

Le nombre de reprises pour un joueur beaucoup plus fort que son adversaire devient

$$\overline{rep} = 1 + \sum_{n=1}^{+\infty} \left(1 - \sum_{i=1}^{n} G_i \right) = 1 + e^{-r} \sum_{n=1}^{+\infty} \frac{r^n}{(n-1)!} \left[\int_0^1 t^{n-1} e^{(1-t)r}\, dt + \frac{n-r-1}{2d} \right] + o\left(\frac{1}{d}\right).$$

En utilisant l'équation (7) ceci donne

$$\overline{rep} = r + 1 + \frac{r^2 e^{-r}}{2d} \left[\sum_{n=2}^{+\infty} \frac{r^{n-2}}{(n-2)!} - \sum_{n=1}^{+\infty} \frac{r^{n-1}}{(n-1)!} \right] + o\left(\frac{1}{d}\right).$$

On retrouve donc

$$\overline{rep} = r + 1 + o\left(\frac{1}{d}\right).$$

Le nombre moyen de reprises est le même à l'ordre 1 qu'à l'ordre 0. Il est vraisemblable que $r+1$ est la valeur exacte, cependant pour le prouver il faudrait refaire les calculs avec l'expression complète de (G_n)...

TAB. 1. Probabilités de jouer la reprise n (c'est-à-dire la probabilité que le match ne soit pas encore fini), de victoire du joueur 1 et de match nul en n reprises.

reprise	probabilité jouer reprise	probabilité match nul	probabilité victoire 1
1ère	1	$G_1(1)G_1(2)$	$G_1(1)\left[1-G_1(2)\right]$
2ème	$\left[1-G_1(1)\right]\left[1-G_1(2)\right]$	$G_2(1)G_2(2)$	$G_2(1)\left[1-G_1(2)-G_2(2)\right]$
...	...	...	...
$n$ième	$\left[1-\sum_{k=1}^{n-1}G_k(1)\right]\left[1-\sum_{k=1}^{n-1}G_k(2)\right]$	$G_n(1)G_n(2)$	$G_n(1)\left[1-\sum_{k=1}^{n}G_k(2)\right]$

2.4. *Déterminer le vainqueur*

G_n est la probabilité que le joueur finisse en n reprises. En comparant les $\{G_n\}$ pour chaque joueur, on peut déterminer la probabilité que chaque joueur gagne. $\sum_{k=1}^{n-1} G_k(J)$ est la probabilité que le joueur J ait atteint d points en $n-1$ reprises. La probabilité de jouer la reprise n est donc la probabilité que 1 n'ait pas fini en $n-1$ reprises fois la probabilité que 2 n'ait pas fini en $n-1$ reprises :

$$\left[1-\sum_{k=1}^{n-1}G_k(1)\right]\left[1-\sum_{k=1}^{n-1}G_k(2)\right].$$

La probabilité de nul à la reprise n est la probabilité que les deux joueurs finissent à cette reprise

$$G_n(1)G_n(2).$$

La probabilité de victoire du joueur 1 à la reprise n est la probabilité que le joueur 1 finisse à la reprise n et que le joueur 2 ne finisse pas pendant les reprises 1 à n :

$$G_n(1)\left[1-\sum_{k=1}^{n}G_k(2)\right].$$

Le tableau 1 donne la probabilité que soit jouée la reprise n et les probabilités d'un match nul et d'une victoire du joueur 1 à la reprise n (la probabilité d'une victoire du joueur 2 s'obtient en permutant les indices 1 et 2.)

2.5. *Calcul de la moyenne*

Le vainqueur marque d points en n reprises, sa moyenne est donc de d/n. Pour ce qui est du perdant (en n reprises), le nombre de points qu'il marque est en moyenne de

$$pts_n^{\mathrm{perd}}=\sum_{i=1}^{d-1} iF_n(i)=(1-p)^n n\sum_{i=1}^{d-1}p^i C_{i+n-1}^{i-1}.$$

La moyenne du joueur 1 est le nombre de points réussis sur le nombre de reprises

$$m^*(1)=\frac{\displaystyle\sum_{n=1}^{+\infty} d\,G_n(1)\left[1-\sum_{k=1}^{n-1}G_k(2)\right]+pts_n^{\mathrm{perd}}\,G_n(2)\left[1-\sum_{k=1}^{n}G_k(1)\right]}{\displaystyle\sum_{n=1}^{+\infty} n\,G_n(1)\left[1-\sum_{k=1}^{n-1}G_k(2)\right]+n\,G_n(2)\left[1-\sum_{k=1}^{n}G_k(1)\right]}. \tag{9}$$

Au numérateur comme au dénominateur le premier terme correspond aux victoires et le second aux défaites.

2.5.1. Convergence

On considère trois cas limites :

(a) Si un joueur est très supérieur à son adversaire, il remportera quasi tous les matches. Sa moyenne calculée est donc de le nombre de points marqués (d) divisé par le nombre de reprises ($\approx \overline{rep}$). Ceci donne

$$\frac{m - m^*}{m} \approx \frac{1}{\overline{rep}}. \tag{10}$$

La moyenne calculée m^* est donc toujours inférieure à la moyenne réelle. En outre la convergence est très lente (en $1/d$). Pour des distances débouchant sur des matches en 2 reprises par exemple, la moyenne du vainqueur est sous-estimée de moitié. Même pour des matches en 10 reprises, il y a 10 % d'erreur.

(b) Dans le cas où les deux adversaires ont le même niveau, la moyenne calculée des deux joueurs est de

$$m^* = \frac{\displaystyle\sum_{n=1}^{+\infty} (d + pts_n^{\text{perd}}) G_n \left[1 - \sum_{k=1}^{n} G_k\right] - dG_n^2}{\displaystyle 2\sum_{n=1}^{+\infty} nG_n \left[1 - \sum_{k=1}^{n} G_k + G_n/2\right]}.$$

Ce cas est plus difficile à calculer.

(c) Si un joueur est beaucoup moins fort que son adversaire, il ne remportera à peu près aucun match. Il peut cependant faire match nul sur de courtes distances. Sa moyenne calculée est donc de

$$m^* = \frac{\displaystyle\sum_{n=1}^{+\infty} pts_n^{\text{perd}} G_n}{\displaystyle\sum_{n=1}^{+\infty} nG_n},$$

où les $\{G_n\}$ sont ceux de son adversaire. Ce cas aussi est difficile à calculer.

2.5.2. En divisant par le nombre de points manqués

Si on divise le nombre de points réussis par le nombre de points manqués au lieu de diviser par le nombre de reprises on obtient

$$m^*_{\text{corrigée}}(1) = \frac{\displaystyle\sum_{n=1}^{+\infty} dG_n(1) \left[1 - \sum_{k=1}^{n-1} G_k(2)\right] + pts_n^{\text{perd}} G_n(2) \left[1 - \sum_{k=1}^{n} G_k(1)\right]}{\displaystyle\sum_{n=1}^{+\infty} (n-1)G_n(1) \left[1 - \sum_{k=1}^{n-1} G_k(2)\right] + nG_n(2) \left[1 - \sum_{k=1}^{n} G_k(1)\right]}, \tag{11}$$

ce qui peut être écrit

$$\frac{m^*_{\text{corrigée}}(1) - m^*(1)}{m^*(1)} = \frac{\displaystyle\sum_{n=1}^{+\infty} G_n(1) \left[1 - \sum_{k=1}^{n-1} G_k(2)\right]}{\displaystyle\sum_{n=1}^{+\infty} (n-1)G_n(1) \left[1 - \sum_{k=1}^{n-1} G_k(2)\right] + nG_n(2) \left[1 - \sum_{k=1}^{n} G_k(1)\right]}.$$

Donc $m^*_{\text{corrigée}}(1) > m^*(1)$. C'est une bonne chose vu que la moyenne telle que calculée par (9) est trop basse. On note que $m^*_{\text{corrigée}}$ est d'autant plus différente de m^* que $\sum_{n=1}^{+\infty} G_n(1) \left[1 - \sum_{k=1}^{n-1} G_k(2)\right]$ est grand. La différence est donc plus grande quand le joueur gagne ou fait match nul, comme on s'y attendait.

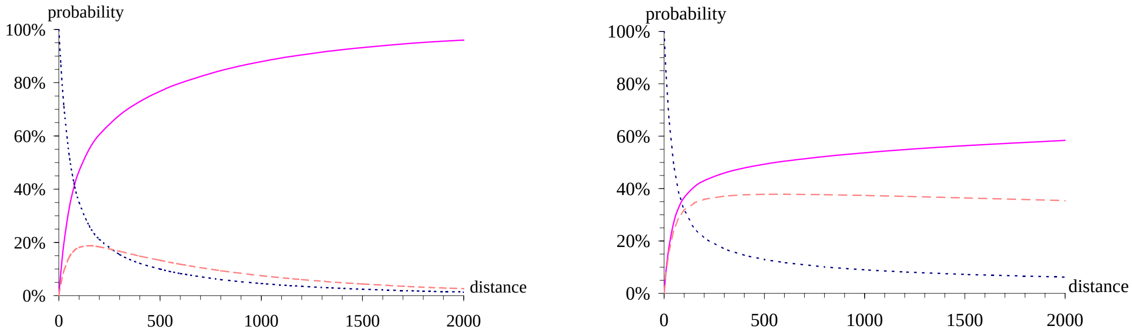

Figure 2. Match opposant un joueur valant (a) 200 de moyenne et (b) 110 de moyenne à un joueur valant 100 de moyenne : probabilités de nul (pointillés courts), de victoire du joueur à 200/110 (ligne continue) et du joueur à 100 (pointillés longs) en fonction de la distance.

En utilisant l'approximation de G_n de l'équation (5), $m^*_{\text{corrigée}} \approx m$ pour un joueur beaucoup plus fort que son adversaire. Ceci montre que $m^*_{\text{corrigée}}$ est une meilleure approximation de la moyenne que m^* et qu'elle est beaucoup moins sensible à la distance.

3. Résultats

On considère deux joueurs de moyennes m_1 et m_2. $\mu_n(1) = p_1^n = \left(\frac{m_1}{1+m_1}\right)^n$. De même $\mu_n(2) = p_2^n = \left(\frac{m_2}{1+m_2}\right)^n$.

3.1. Joueurs à 110 et 200 de moyenne contre un joueur à 100 de moyenne

La figure 2(a) donne le pourcentage de victoires de chaque joueur et le pourcentage de matches nuls en fonction de la distance du match pour un match opposant un joueur à 200 de moyenne à un joueur à 100 de moyenne. La probabilité de victoire du joueur à 100 de moyenne varie assez peu pour des matches entre 60 et 400 points. Malgré l'écart de niveau important entre les deux joueurs, sur des distances inférieures à 400 points le joueur le plus faible a plus de 15 % de chances de gagner. Le joueur le plus fort gagne à peine les deux tiers des matches pour des matches en 300 points. Sur des distances nettement plus longues le meilleur gagne presqu'à chaque fois.

La figure 2(b) donne le pourcentage de victoires de chaque joueur et le pourcentage de matches nuls en fonction de la distance du match pour un match opposant un joueur à 110 de moyenne à un joueur à 100 de moyenne. Sur des distances supérieures à 300 points, le joueur à 100 de MG gagne environ un tiers des matches, ce chiffre varie peu en fonction de la distance, ce qui est assez inattendu. Pour des joueurs de niveaux proches la probabilité de victoire du joueur le plus fort n'est pas beaucoup plus élevée que celle de son adversaire même pour des marathons en 1 000 ou 2 000 points.

3.1.1. Conséquences

Au niveau européen ou mondial (ce qui revient au même) les parties de cadre 47/2 en 250 ou 300 points sont inadaptées. Il faudrait :

(a) instaurer des parties nettement plus longues (400 ou 500 points par exemple) lors des phases finales (ceci n'est pas impossible, des parties de libre en 2 000 ou 3 000 points ont existé par le passé) ou
(b) arrêter d'organiser des championnats de 47/2 comme ont été abandonnées les compétitions de libre à ce niveau ou
(c) modifier les règles pour empêcher les joueurs d'avoir des séries de 200 ou 300 points ‡.

‡Jerry Briesath a inventé une variation du 14.1 continu appelée "equal offense" où un joueur ne peut marquer plus de 20 points d'affilée. Au billard carambole, si un joueur a une série de 50 par exemple, on remettrait sur

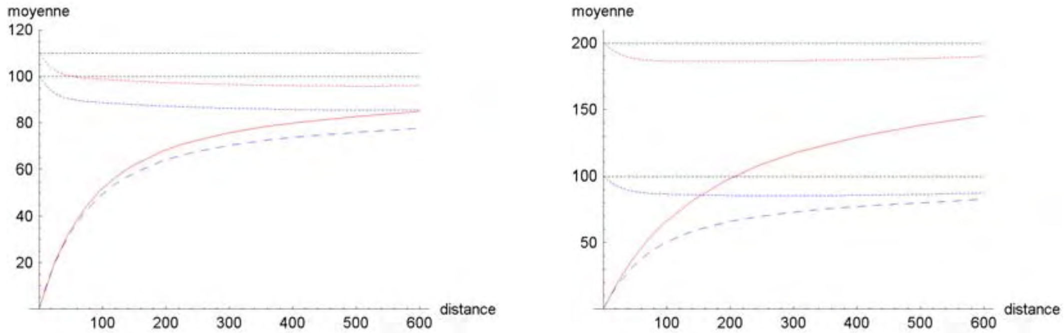

Figure 3. Moyenne des deux adversaires en fonction de la distance. Les lignes en pointillés courts sont des moyennes corrigées où on divise par le nombre de points manqués. (a) Match 200-100 et (b) match 110-100.

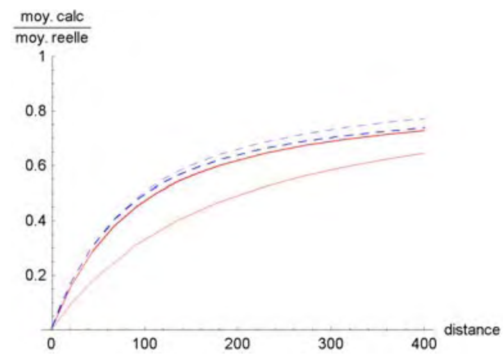

Figure 4. Moyenne des joueurs divisée par leur moyenne théorique en fonction de la distance. Les lignes continues sont pour le meilleur joueur (110 ou 200 de MG) et les lignes pointillées pour le joueur à 100 de moyenne, les lignes fines correspondent aux matches 200-100 et les lignes épaisses aux matches 110-100.

3.2. Moyenne

Comme le montre la figure 3, si on calcule la moyenne en divisant le nombre de points marqués par le nombre de reprises le résultat est catastrophique. Le joueur qui vaut 200 de moyenne a l'air de n'avoir une moyenne que de 120 sur un match en 300 points. Comme le note J.-M. Fray (2002), ceci est dû à ce qu'on considère que la dernière reprise du vainqueur équivaut à un point manqué. Si un joueur valant plus de 100 de MG joue une partie en 100 points, ou bien il finit en une reprise et a une moyenne de 100 ou bien il finit en plus d'une reprise et aura une moyenne d'au mieux 50. Dans le premier cas il n'a pas raté un seul point et dans le second il a marqué 100 points en manquant une seule fois. Si on divise le nombre de points marqués par le nombre de points manqués (équation (11)) on obtient un résultat fort différent (lignes en pointillées courts sur la figure 3). Les moyennes "corrigées" sont meilleures que la formule traditionnelle, cependant elles sont toujours inférieures à la moyenne réelle. Ceci est un peu surprenant mais il n'y a pas d'incohérence, les moyennes "corrigées" sont de meilleures approximations de la moyenne réelle même si elles ne sont pas la panacée qu'on aurait pu imaginer.

La figure 4 donne, en fonction de la distance, le ratio de la moyenne calculée sur le match sur la moyenne théorique. Le résultat le plus mauvais est obtenu par le joueur à 200 de MG contre le joueur à 100 de MG, le joueur à 100 de MG est toujours celui dont la moyenne est la moins sous-estimée. Si la moyenne est utilisée pour classer les joueurs, un joueur a donc intérêt à perdre ! Pour cela choisir des adversaires plus forts que soi. Les joueurs qui veulent ou doivent gagner auront intérêt à jouer sur de longues distances.

mouches et son adversaire jouerait jusqu'à ce qu'il manque ou atteigne 50.

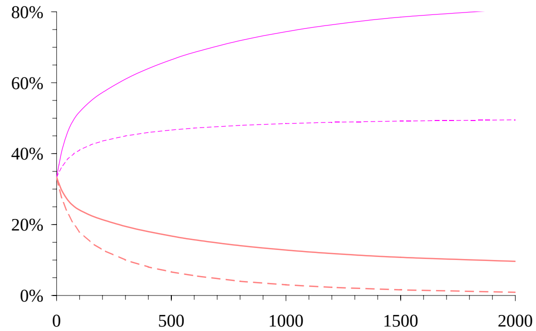

Figure 5. Probabilités de sortie de poule (en fonction de la distance des matches) dans le cas d'un joueur valant 200 de moyenne opposé à deux joueurs à 100 (lignes continues) et de deux joueurs à 200 opposés à un joueur valant 100 de moyenne (lignes pointillées). Les lignes fines correspondent aux probabilités de sortie de poule du (des) joueur(s) à 200 de moyenne et les lignes épaisses aux probabilités du (des) joueur(s) à 100.

4. Conclusion

On a montré que le nombre moyen de reprises est de l'ordre de la distance divisée par la moyenne du joueur. Les distances peu supérieures à la moyenne des joueurs donnent le plus fréquemment des matches en deux reprises. Au niveau mondial les parties de cadre 47/2 en 250 ou 300 points sont donc trop courtes. Si on calcule la moyenne en divisant le nombre de points réussis par le nombre de points manqués (au lieu de diviser par le nombre de reprises), le résultat est assez proche de la moyenne théorique même pour de très courtes distances. Mais ensuite la convergence est très lente : la moyenne n'est guère mieux évaluée par des parties en 400 points que par des parties en 100 points. Cette façon de calculer la moyenne est donc meilleure mais pas autant qu'on pourrait le penser.

Annexe 1 - Poules

On considère une poule de trois joueurs dont deux de même niveau. Par exemple une tête de série et deux joueurs de moindre niveau (joueurs "chapeau"). On appelle p_V la probabilité de victoire de la tête de série dans ses matches contre les chapeaux, p_N est la probabilité d'un nul et p_D la probabilité d'une défaite de la tête de série. Pour le match entre les deux chapeaux les probabilités sont p'_V, p'_N et p'_D. On notera que ces joueurs ayant le même niveau, $p'_D = p'_V$. Si un seul joueur sort de poule, la probabilité que ce soit la tête de série est de

$$p_V + p_V \left[(p_N - p_D) \left(1 - \frac{2}{3} p'_V \right) - \frac{p_N p'_V}{3} \right] - \frac{p'_N p_N^2}{3}.$$

La figure 5 représente (en fonction de la distance des matches) la probabilité de sortie de poule d'un joueur à 200 de MG opposé à deux joueurs à 100 de MG. Dans le cas de matches en 300 points, la tête de série a environ 60 % de chances de sortir de poule.

On a considéré le cas d'une tête de série avec deux joueurs "chapeau". On peut aussi s'intéresser au cas de deux têtes de série se retrouvant dans la même poule (*cf.* déboires de J. Reverchon l'année dernière). Les lignes pointillées de la figure 5 représentent la probabilité de sortie de poule de deux joueurs à 200 de MG opposés à un joueur à 100 de MG. Si les matches sont en 300 points, les "têtes de série" ont environ 45 % de chances de sortir de poule, le troisième joueur a à peine 10 % de chances.

Annexe 2 - Calcul des moyennes générale et glissante

La moyenne sur plusieurs matches (moyenne générale ou glissante) est calculée comme la somme des points marqués divisée par la somme des reprises

$$MG = \frac{\sum_{k=1}^{N} d_k}{\sum_{i=1}^{N} n_i}.$$

N est le nombre de matches. La moyenne générale (MG) n'est pas la moyenne des moyennes des matches (ici et dans ce qui suit on pourra substituer moyenne glissante à la place de moyenne générale vu que cette annexe traite des moyennes sur plusieurs matches en général). Par exemple, si les moyennes sur trois matches sont 1/2, 1/2 et 2, les nombres de reprises sont (pour des matches en 40 points) 80, 80 et 20. La moyenne générale est donc de $120/180 = 2/3$. La moyenne des moyennes est de $3/3 = 1$.

L'inverse de la moyenne est

$$\frac{1}{MG} = \frac{\sum_{i=1}^{N} n_i}{\sum_{k=1}^{N} d_k} = \sum_{i=1}^{N} \frac{1}{m_i} \frac{d_i}{\sum_{k=1}^{N} d_k} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{m_i} \frac{d_i}{\langle d \rangle}.$$

$\langle d \rangle$ est la distance moyenne. Si tous les $\{d_i\}$ sont égaux (tous les matches sur la même distance) alors $d_i = \langle d \rangle \; \forall \, i$ et

$$\frac{1}{MG} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{m_i}.$$

Dans ce cas, la moyenne est la moyenne inverse des moyennes. En reprenant l'exemple de trois matches à 1/2, 1/2 et 2 de moyenne, les inverses des moyennes sont 2, 2 et 0,5. L'inverse de la moyenne générale est la moyenne de ces inverses : $4,5/3 = 3/2$. On obtient que la moyenne générale vaut 2/3, ce qui est conforme à ce qu'on a calculé précédemment.

Si les deux matches à 0,5 sont des défaites où le joueur n'a marqué que 25 points alors que le match à 2 de moyenne est une vocitoire (en 40 points) alors la MG vaut $90/120 = 3/4$. La moyenne est plus élevée que dans le cas précédent (2/3) parce que le match à 2 de moyenne contribue plus (plus de points marqués sur ce match). Dans ce cas la moyenne inverse des moyennes n'est qu'une approximation de la MG. Cette approximation est néanmoins meilleure que la bête moyenne des moyennes.

En general tous les matches ne sont pas sur la même distance. L'erreur introduite par la moyenne inverse est

$$\frac{1}{MG} - \frac{1}{N} \sum_{i=1}^{N} \frac{1}{m_i} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{m_i} \frac{d_i - \langle d \rangle}{\langle d \rangle}.$$

$$MG - \frac{1}{N} \sum_{i=1}^{N} m_i = \frac{1}{N} \sum_{k=1}^{N} d_k \left(\frac{1}{\langle n \rangle} - \frac{1}{n_k} \right).$$

On a donc montré que si le joueur marque le même nombre de points sur tous ses matches, sa moyenne générale peut bel et bien se calculer à partir des moyennes par match, mais pas en utilisant la moyenne arithmétique : la MG est la moyenne inverse des moyennes. Donc si par fainéantise ou manque d'information on ne veut (ou ne peut) pas calculer la MG en utilisant les nombres de points marqués et de reprises, on peut calculer la moyenne inverse des moyennes. Le résultat n'est exact que si tous les matches sont en un même nombre de points, mais ce résultat est quand même une meilleure approximation que la simple moyenne des moyennes dans tous les cas.

Bibliographie

Bouville, M. (2005a). Le replacement au billard carambole : un processus de Markov. (disponible à : http ://arxiv.org/pdf/math.PR/0508089. La version française suit la version anglaise).

Bouville, M. (2005b). Le suisse, le russe et l'américaine. (disponible à http ://billiards.mathieu.bouville.name/suisse-russe-americaine.pdf).

Fray, J.-M. (2002). A propos de la moyenne au billard carambole. *in* F. Caudron, *Le billard en expansion.* Deleye, Ledegem, Belgium.