Tennis et propagation de biais

Angelo ROSELLO

Plan

Issue d'un jeu au tennis : description du modèle et résultats
- Description du modèle
- Probabilité de victoire
- Incertitude et entropie
- Ouverture

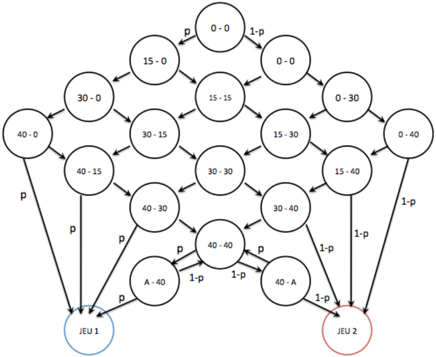
Etude statistique et validation du modèle
- ATP World Tour Final 2011
- US Open Final 2013

Description du modèle : chaine de Markov et biais

"Biais" $p \in [0,1]$: probabilité que le joueur 1 marque un point (sur une balle)

- $p > \frac{1}{2}$: biais en faveur du joueur 1
- $p < \frac{1}{2}$: biais en faveur du joueur 2
- $p = \frac{1}{2}$: pas de biais, partie indécidable

Question : Quelle est la probabilité $J(p)$ que le joueur 1 remporte le jeu ?



Probabilité de victoire : calcul

On fixe un biais $p \in [0, 1]$.

- Notation : Pour un score s, on désigne par $P_J(s)$ la probabilité que le joueur 1 remporte le jeu sachant que le score actuel est s.
- Objectif : Calculer $P_J(0 \text{ - } 0) = J(p)$.
- Propriété de Markov : A un temps donné, le processus est indépendant de son historique ("Le futur ne dépend du passé que par le présent").
- En partant du score (0 - 0), le prochain score est (15 - 0) avec probabilité p, ou (0 - 15) avec probabilité $1 - p$. La propriété de Markov donne :

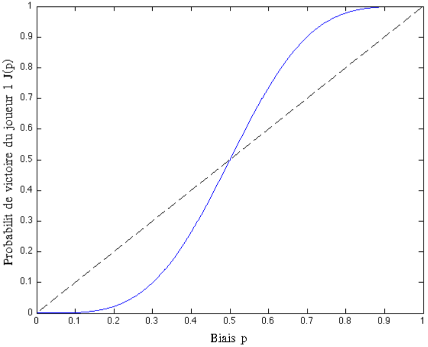
$$P_J(0 \text{ - } 0) = p \times P_J(15 \text{ - } 0) + (1 - p) \times P_J(0 \text{ - } 15) \tag{1}$$

  On poursuit similairement les calculs de $P_J(15 \text{ - } 0)$ et $P_J(0 \text{ - } 15)$ en "descendant" dans le graphe.
- Le calcul aboutit car les probabilité "extrémales" $P_J(\text{JEU } 1) = 1$ et $P_J(\text{JEU } 2) = 0$ sont connues. Seul le calcul de $P_J(40 \text{ - } 40)$ diffère un peu du reste (boucle).

Probabilité de victoire : résultat

Au final, on obtient :

$$\boxed{J(p) = p^4 + 4p^4(1-p) + 10p^4(1-p)^2 + \frac{20p^5(1-p)^3}{1-2p(1-p)}} \quad (2)$$



Le biais initial p (sur une balle) est accentué :

Si $p = \frac{3}{4}$ (3 balles sur 4 remportées en moyenne), $J(p) \simeq 0.9492$
(95% de chance de remporter le jeu)

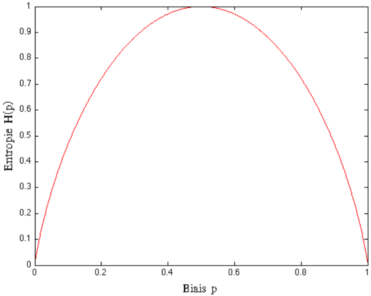
Si $p = \frac{1}{4}$ (1 balle sur 4 remportée en moyenne), $J(p) \simeq 0.0508$
(5% de chance de remporter le jeu)

Incertitude et entropie : entropie d'un échange

De manière générale, une entropie (S en physique, souvent H en maths) est une fonction qui mesure la quantité d'information, le désordre, l'incertitude d'un système. Intuitivement, plus un système est complexe et imprévisible, plus son entropie est élevée.

L'incertitude quand à l'issue d'un échange de biais p est ici quantifiée par l'*entropie binaire* (entropie de Shannon) :

$$\boxed{H(p) = -p\log_2(p) - (1-p)\log_2(1-p)} \tag{3}$$



$p = 0 \Rightarrow$ Joueur 2 gagne
$\rightarrow$ Issue certaine : $H(p) = 0$

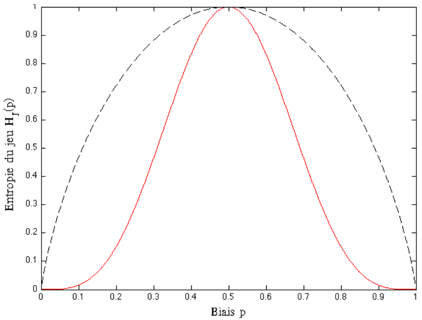
$p = 1 \Rightarrow$ Joueur 1 gagne
$\rightarrow$ Issue certaine : $H(p) = 0$

$p = \frac{1}{2} \Rightarrow$ Pas de biais
$\rightarrow$ Indécidable : $H(p) = 1$ (incertitude maximale)

Incertitude et entropie : entropie du jeu

Le biais p sur un échange induit, comme on l'a vu, un biais $J(p)$ "accentué" sur l'issue du jeu. L'incertitude quand à l'issue du jeu dans sa totalité est donc quantifiée par :

$$\boxed{H_J(p) = H(J(p)) = -p\log_2(J(p)) - (1 - J(p))\log_2(1 - J(p))} \tag{4}$$



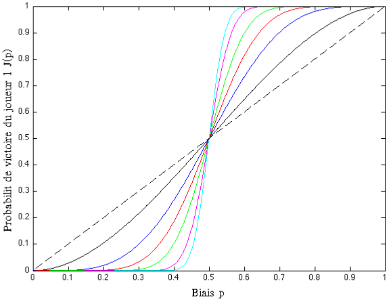
L'incertitude "se dissipe" en considérant le jeu dans sa totalité :

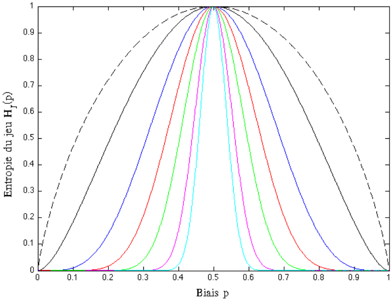
Bien que le joueur 1 soit par exemple meilleur que le joueur 2 ($p > 1/2$), il est difficile de prévoir l'issue d'un seul échange (incertitude $H(p)$ élevée). L'issue du jeu est en revanche plus facilement prévisible (incertitude $H_J(p)$ plus faible).

Lorsque $p = \frac{1}{2}$, l'issue du jeu reste indécidable : $H_J(p) = 1$. (incertitude maximale)

Ouverture

On comprend que, du fait qu'un jeu consiste en une succession d'échanges de biais constant $p \neq \frac{1}{2}$ (cas décidable), le biais final sur la totalité du jeu $J(p)$ est fortement accentué.
En s'écartant du cadre du tennis, on pourrait considérer un jeu qui consiste en une succession de parties de biais constant $p \in [0, 1]$, le premier joueur atteignant N parties gagnantes remportant le jeu.
Lorsque $N \to \infty$, on observe qu'un biais "insignifiant" sur une partie résulte en un biais important sur l'issue du jeu.





Etude statistique : ATP World Tour Final 2011

Roger Federer

Echanges gagnés : 105

Jo-Wilfried Tsonga

Echanges gagnés : 91

Biais empirique : $\boxed{\hat{p} = \frac{105}{105+91} \simeq 0.5357 \simeq 54\%}$ en faveur de Federer.

Le modèle prévoit donc que Federer gagne en moyenne $\boxed{J(\hat{p}) \simeq 0.5886 \simeq 59\%}$ des jeux.

En effet, il gagne 18 jeux sur 31, soit $\frac{18}{31} \simeq 0.5806 \simeq 58\%$ des jeux !

Etude statistique : US Open Final 2013

Novak Djokovic	Rafael Nadal
	
Echanges gagnés : 101	Echanges gagnés : 121

Biais empirique : $\boxed{\hat{p} = \dfrac{101}{101+121} \simeq 0.4549 \simeq 45\%}$ en faveur de Nadal.

Le modèle prévoit donc que Djokovic gagne en moyenne $\boxed{J(\hat{p}) \simeq 0.3887 \simeq 39\%}$ des jeux.

En effet, il gagne 13 jeux sur 34, soit $\frac{13}{34} \simeq 0.38235 \simeq 38\%$ des jeux !

$\boxed{\text{Modèle validé ! :)}}$