Linear Reward Inaction

Claire BASKEVITCH,Tristan BESSAC,
Joseph DESQUAIRES, Bin LIU

Paris Saclay

Février 2020

Sommaire

Sommaire

Sommaire

Sommaire

Sommaire

Le jeu

Il s'agit d'un jeu de poker simplifié où les 2 joueurs tirent une carte entre 1 et 9 en même temps.

Alice joue en première - 4 choix :

- Se coucher.
- Relancer de 1.
- Relancer de 2.
- Relancer de 4.

Bob joue toujours en deuxième - 2 choix :

- Suivre la mise de Alice.
- Se coucher.

Sommaire

Principe

- Repose sur un **principe de récompense** lors d'une action positive.
- **Ne prends en compte ni la perte ni l'égalité**.

- On a un **vecteur stochastique de stratégie** pour chaque carte et joueur qui **se met à jour uniquement lors d'un gain**.
- Permet de trouver un **équilibre de Nash (en stratégies pures)** pour un.

Algorithme de LRI

Données : $b \in [0, 1]$ un réel
Initialisation : $\forall \ell \in N, s \in \{1, \ldots, K\}, \; P_{s,\ell}(1) = \frac{1}{K}$
Pour chaque *itération* t **faire**
  Pour chaque *joueur* ℓ **faire**
    Tirer une stratégie s aléatoirement en respectant le vecteur de probabilités $P_\ell(t)$
  Fin
  Pour chaque *joueur* ℓ **faire**
    Recevoir un gain $u_\ell(S(t)) \in [0, 1]$
    Mettre à jour les probabilités des stratégies selon la règle suivante
  Fin
Fin

K : nombre de stratégies.

Mise à jour des stratégies des joueurs

Règle de Mise à jour

$$q_{i,s}(t+1) = \begin{cases} q_{i,s}(t) + b * U_t * (1 - q_{i,s}(t)) & \text{Si } s = s_i(t) \\ q_{i,s}(t) - b * U_t * q_{i,s}(t) & \text{Si } s \neq s_i(t) \end{cases}$$

Variables et constantes

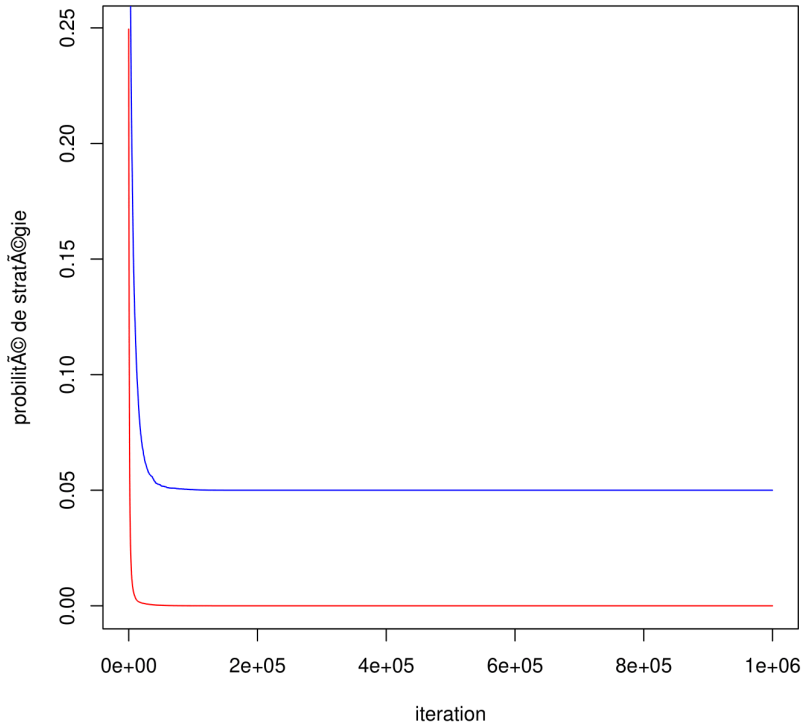
b : paramètre d'apprentissage tq $b \in [0, 1]$ et $b \leq \frac{1}{U_{max}}$.

$q_{i,s}(t)$: probabilité que le joueur i joue la stratégie s à l'étape t.

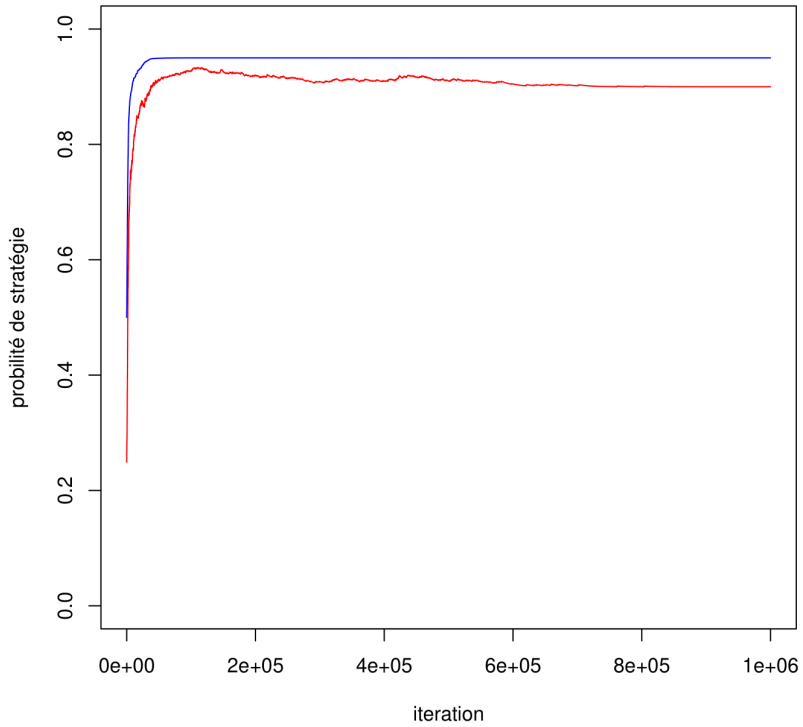
U_t : fonction d'utilité.

Étude de courbes

Probabilité qu'Alice se couche
Probabilité que Bob se couche

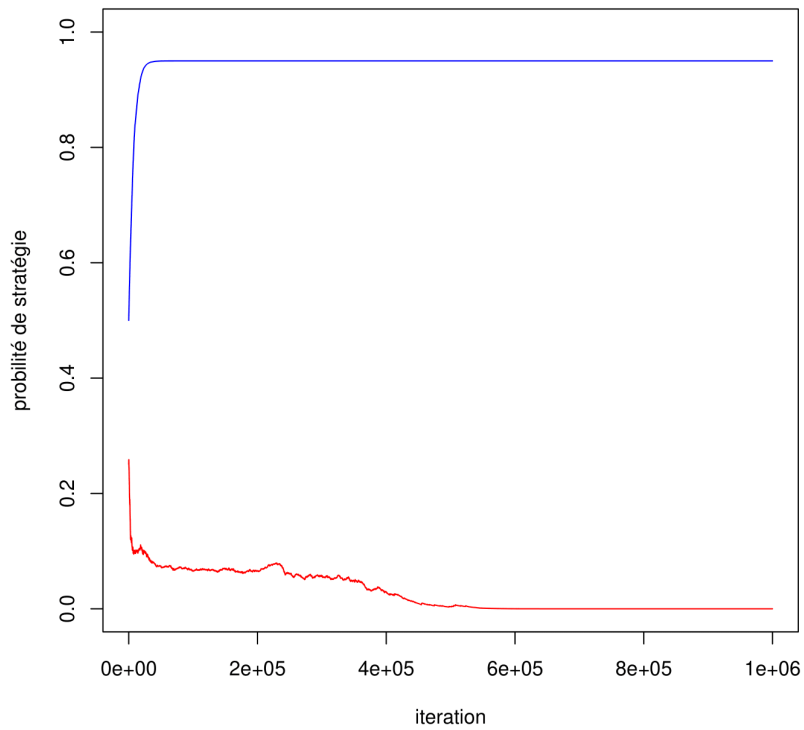


Probabilité qu'Alice relance de 4
Probabilité que Bob suive

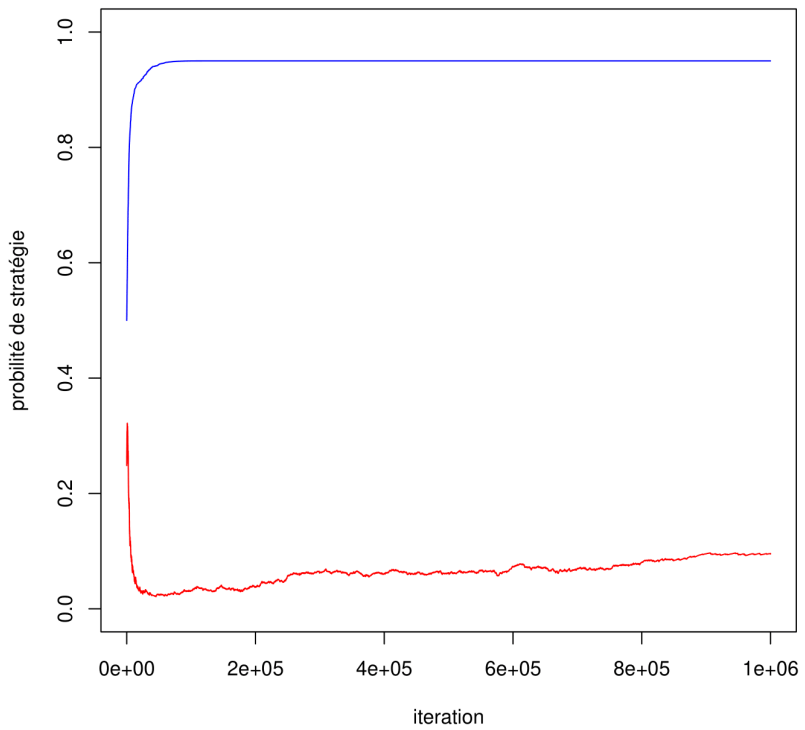


Étude de courbes

Probabilité qu'Alice relance de 1
Probabilité que Bob suive



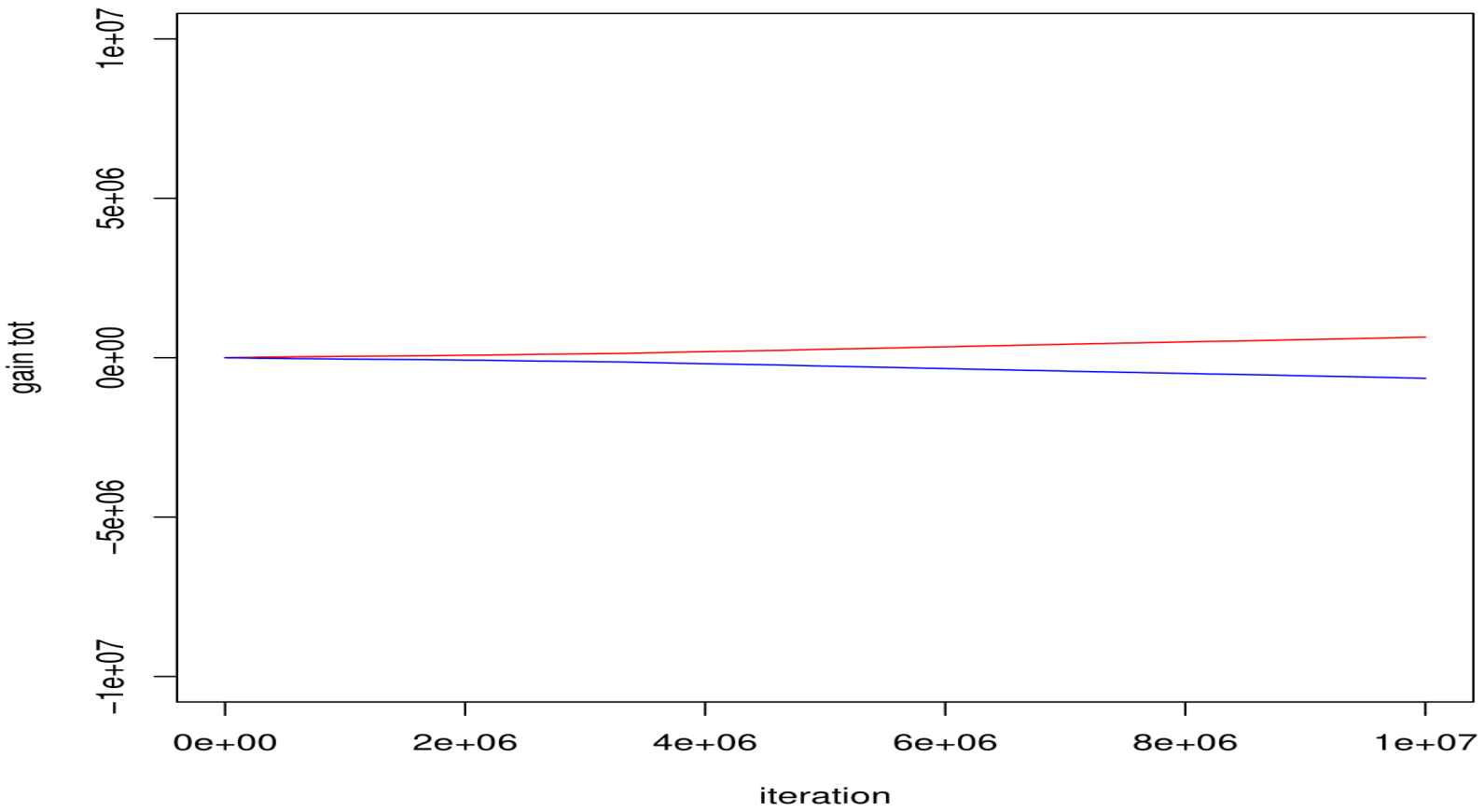
Probabilité qu'Alice relance de 2
Probabilité que Bob suive



université PARIS-SACLAY

Étude de courbes

Courbes des gains de Bob et dAlice .



Sommaire

Principe

- Repose sur un **principe de récompense** lors d'une action positive.
- **Pénalise** les mauvaises actions.

- On obtient un **vecteur stochastique de stratégies** pour chaque carte et joueurs qui **se met à jour lors d'un gain ou d'une perte**.

Mise à jour des stratégies des joueurs

Règle de Mise à jour

$$q_{i,s}(t+1) = \begin{cases} q_{i,s}(t) + b * U_t * (1 - q_{i,s}(t)) - \beta * q_{i,s}(t) * (1 - U_t) & \text{Si } s = s_i(t) \\ q_{i,s}(t) - b * U_t * q_{i,s}(t) + \beta * ((k-1)^{-1} - q_{i,s}(t)) * (1 - U_t) & \text{Si } s \neq s_i(t) \end{cases}$$

Variables et constantes

- b : paramètre d'apprentissage tq $b \in [0,1]$ et $b \leq \frac{1}{U_{max}}$.
- β : paramètre d'apprentissage tq $\beta \in [0,1]$
- k : nombre total des actions.
- $q_{i,s}(t)$: probabilité que le joueur i joue la stratégie s à l'étape t.
- U_t : fonction d'utilité.

Mise à jour des stratégies des joueurs

Règle de Mise à jour

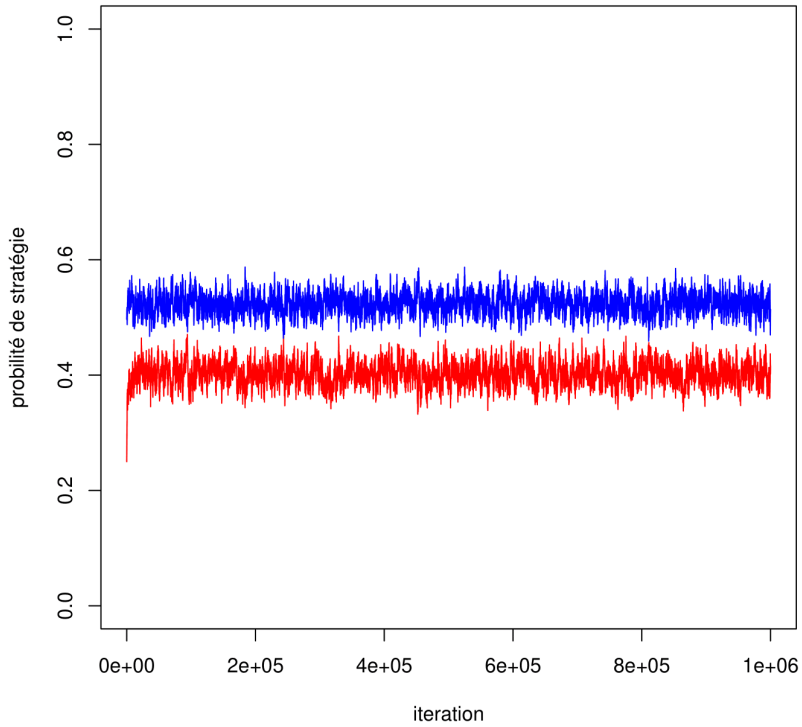
$$q_{i,s}(t+1) = \begin{cases} q_{i,s}(t) + b * U_t * (1 - q_{i,s}(t)) - \beta * q_{i,s}(t) * (1 - U_t) & \text{Si } s = s_i(t) \\ q_{i,s}(t) - b * U_t * q_{i,s}(t) + \beta * ((k-1)^{-1} - q_{i,s}(t)) * (1 - U_t) & \text{Si } s \neq s_i(t) \end{cases}$$

Décision du choix des règles

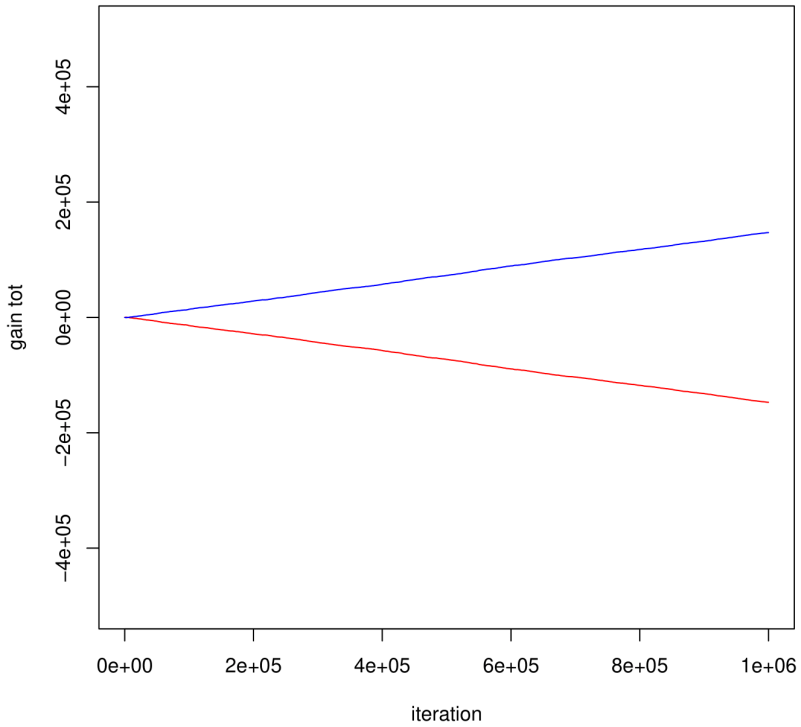
$$\begin{cases} \textit{Linear Reward Penalty} & \text{Si } b = \beta \\ \textit{Linear Reward Inaction} & \text{Si } \beta = 0 \\ \textit{Linear Reward } \varepsilon - \textit{Penalty} & \text{Si } b >> \beta \end{cases}$$

Etude de courbes

Probabilité qu'Alice relance de 4
Probabilité que Bob suive



Gain total d'Alice et de Bob



Sommaire

Conclusion

- LRI est peu efficace sur des jeux avec perte.
- LRP est donc plus intéressant sur ce jeu.