Preuve ε-greedy

Dilemme du prisonnier

Dans le dilemme du prisonnier, on a la matrice de payoff suivante:

prisonnier A \ B	cooperates	defects
cooperates	(R, R)	(S, T)
defects	(T, S)	(P, P)

T: Temptation
R: Reward
P: Punishment
S: Sucker's

$$\boxed{T > R > P > S}$$

Par exemple

A\B	C	D
C	(-1, -1)	(-3, 0)
D	(0, -3)	(-2, -2)

+ À chaque pas de temps, les agents A et B vont jouer une action (ex (C, C))

Après n pas de temps

$N[(C,C)] = a$ avec $a+b+c+d=n$.
$N[(C,D)] = b$
$N[(D,C)] = c$
$N[(D,D)] = d$

Après n essais, le nombre de possibilités pour $a,b,c,d \in \mathbb{N}$ tel que $a+b+c+d=n$ est donné par $\binom{n+3}{3}$

Si on considère $a,b,c,d \in \mathbb{N} \setminus \{0\}$, on obtient $\binom{n-1}{3}$ possibilités

On veut partionner n symboles dans 4 groupes, donc on a besoin de 3 séparateurs.

$$\underbrace{\circ\circ\circ \mid \circ\circ\circ \mid \circ\circ\circ\circ}_{n+3 \text{ symboles}}$$

On choisit les emplacements possibles de 3 séparateurs dans $n+3$ positions possibles.

Définition des moyennes après n pas de temps

Rappel: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $\begin{bmatrix} (R,R) & (S,T) \\ (T,S) & (P,P) \end{bmatrix}$

Pour l'agent A:

$$\mu_{AC} = [a \cdot R + b \cdot S] / (a+b) = S + a/a+b \, (R-S)$$
$$\mu_{AD} = [c \cdot T + d \cdot P] / (c+d) = P + c/(c+d) \, (T-P)$$

Pour l'agent B:

$$\mu_{BC} = [a \cdot R + c \cdot S] / (a+c) = S + a/a+c \, [R-S]$$
$$\mu_{BD} = [b \cdot T + d \cdot P] / (b+d) = P + d/b+d \, (T-P]$$

Supposons qu'après n pas de temps,

$\mu_{AC} > \mu_{AD}$ et $\mu_{BC} > \mu_{BD}$

Les agents coopèrent et sont dans la situation qui maximise leurs récompenses moyennes.

Je vais montrer que si les deux agents suivent une politique ε-greedy, ils vont sortir de cette sitation presque surement

Si $\mu_{AC} > \mu_{AD}$ et $\mu_{BC} > \mu_{BD}$

La probabilité au temps $n+1$ d'avoir (C,C) est donné par

$$P_{n+1}((C,C)) = (1-\varepsilon)^2$$

Si les agents suivent une politique ε'-greedy, $\varepsilon' = 2\varepsilon$

On va donc obtenir les probabilités suivantes:

A\B	C	D
C	$(1-\varepsilon)^2$	$\varepsilon(1-\varepsilon)$
D	$(1-\varepsilon)\varepsilon$	ε^2

Loi forte des grands nombre

Si $(X_n)_{n\geq 0}$ est une suite de variables aléatoires I.I.D

$E(|X_1|) < \infty \Leftrightarrow$ la suite $\frac{X_1 + \dots + X_n}{n}$ Converge presque sûrement

et si une des deux conditions est remplie alors la suite $\frac{X_1 + \dots + X_n}{n} \xrightarrow{p.s} E[X_1]$

Dans notre situation

$$P_a = (1-\varepsilon)^2 \Rightarrow \frac{a_n}{n} \xrightarrow{p.s} (1-\varepsilon)^2$$

$$P_b = \varepsilon(1-\varepsilon)^2 \Rightarrow \frac{b_n}{n} \xrightarrow{p.s} \varepsilon(1-\varepsilon)$$

$$P_d = \varepsilon^2 \Rightarrow \frac{d_n}{n} \xrightarrow{p.s} \varepsilon^2$$

Donc $\frac{a_n}{a_n+b_n} \xrightarrow{p.s} \frac{(1-\varepsilon)^2}{(1-\varepsilon)^2+(1-\varepsilon)\varepsilon} = \frac{1-\varepsilon}{1-\varepsilon+\varepsilon} = 1-\varepsilon$

$$\frac{c_n}{c_n+d_n} \xrightarrow{p.s} \frac{\varepsilon(1-\varepsilon)}{\varepsilon(1-\varepsilon)+\varepsilon^2} = \frac{1-\varepsilon}{1-\varepsilon+\varepsilon} = 1-\varepsilon$$

En remplaçant dans les équations du dilemme du prisonnier

$$\mu_{AC} = S+(1-\varepsilon)(R-S) = \varepsilon S+(1-\varepsilon)R$$
$$\mu_{AD} = P+(1-\varepsilon)(T-P) = \varepsilon \hat{P}+(1-\varepsilon)\hat{T}$$

Or comme $T > R > P > S$, on a que $\mu_{AC} < \mu_{AD}$,

Ce qui est contraire à la situation initiale.

On a donc prouver que ε-greedy fait sortir de la position de collusion.

À quelle vitesse est-ce que les agents sortent de
$\mu_{AC} > \mu_{AD}$ et $\mu_{BC} > \mu_{BD}$

Je vais commencer en simplifiant le problème et approximer la situation ~~les~~ updates comme qqch de continue et déterministe.

En considérant seulement le prisonnier A

$$\mu_{AC} = S + \frac{a + (1-\varepsilon)^2 X}{a+b + [(1-\varepsilon)^2 + \varepsilon(1-\varepsilon)]X} (R-S)$$

$$\mu_{AD} = P + \frac{c + \varepsilon(1-\varepsilon)X}{c+d + [\varepsilon(1-\varepsilon) + \varepsilon^2]X} (T-P)$$

Pour Trouver le moment pour lequel la situation change:

$\mu_{AC} - \mu_{AD} = 0$

$E[X]$

Inégalités sur les counts

$$\frac{a \cdot R + bS}{a+b} - \frac{cT + dP}{c+d} > 0$$

$$(a \cdot R + bS)(c+d) - (cT + dP)(a+b) > 0$$

$$acR + adR + bcS + bdS - acT - bcT - adP - bdP > 0$$

$$ac(R-T) + bd(S-P) + ad(R-P) + bc(S-T) > 0$$

$$a\,[c(R-T) + d(R-P)] + b\,[d(S-P) + c(S-T)] > 0$$

$\Delta_3 \ \Delta_2 \ \Delta_1$

$T > R > P > S$

$3 > 2 > 1 > 0$

$$\frac{a}{b} > \frac{d(P-S) + c(T-S)}{d(R-P) + c(R-T)}$$

$$\frac{a}{b} > \frac{d\Delta_1 + c(\Delta_1 + \Delta_2 + \Delta_3)}{d\Delta_2 + c(\Delta_3)}$$

$$r_1 > \frac{\Delta_1 + r_2(\Delta_1 + \Delta_2 + \Delta_3)}{\Delta_2 + r_2\Delta_3}$$

$$\boxed{\frac{d}{dr_2} = \frac{(\Delta_1 + \Delta_2 + \Delta_3)\Delta_2 - \Delta_1\Delta_3}{(\Delta_2 + r_2\Delta_3)^2} > 0 \qquad \Delta_2^2 + \Delta_1\Delta_2 + \Delta_3\Delta_2 - \Delta_1\Delta_3 > 0}$$

$$\frac{1+3x}{1+x}$$

>0 <0

$$r_1 > \frac{\Delta_1 + \Delta_2 + \Delta_3}{\Delta_3}$$

$$r_1 > \frac{\Delta_1}{\Delta_2}$$