❒ **Epsilon-glouton** – La politique epsilon-gloutonne (en anglais *epsilon-greedy*) est un algorithme essayant de trouver un compromis entre l'exploration avec probabilité ϵ et l'exploitation avec probabilité $1-\epsilon$. Pour un état s, la politique π_{act} est calculée par :

$$\pi_{\text{act}}(s) = \begin{cases} \underset{a\in \text{ Actions}}{\text{argmax}} \ \widehat{Q}_{\text{opt}}(s,a) & \text{avec proba } 1-\epsilon \\ \text{random from Actions}(s) & \text{avec proba } \epsilon \end{cases}$$

2.3 Jeux

Dans les jeux (e.g. échecs, backgammon, Go), d'autres agents sont présents et doivent être pris en compte au moment d'élaborer une politique.

❒ **Arbre de jeu** – Un arbre de jeu est un arbre détaillant toutes les issues possibles d'un jeu. En particulier, chaque nœud représente un point de décision pour un joueur et chaque chemin liant la racine à une des feuilles traduit une possible instance du jeu.

❒ **Jeu à somme nulle à deux joueurs** – C'est un type jeu où chaque état est entièrement observé et où les joueurs jouent de manière successive. On le définit par :

- un état de départ s_{start}
- de possibles actions Actions(s) partant de l'état s
- du successeur Succ(s,a) de l'état s après avoir effectué l'action a
- la connaissance d'avoir atteint ou non un état final IsEnd(s)
- l'utilité de l'agent Utility(s) à l'état final s
- le joueur Player(s) qui contrôle l'état s

Remarque : nous assumerons que l'utilité de l'agent a le signe opposé de celui de son adversaire.

❒ **Types de politiques** – Il y a deux types de politiques :

- Les politiques déterministes, notées $\pi_p(s)$, qui représentent pour tout s l'action que le joueur p prend dans l'état s.
- Les politiques stochastiques, notées $\pi_p(s,a) \in [0,1]$, qui sont décrites pour tout s et a par la probabilité que le joueur p prenne l'action a dans l'état s.

❒ **Expectimax** – Pour un état donné s, la valeur d'expectimax $V_{\text{exptmax}}(s)$ est l'utilité maximum sur l'ensemble des politiques utilisées par l'agent lorsque celui-ci joue avec un adversaire de politique connue π_{opp}. Cette valeur est calculée de la manière suivante :

$$V_{\text{exptmax}}(s) = \begin{cases} \text{Utility}(s) & \text{IsEnd}(s) \\ \underset{a\in\text{Actions}(s)}{\max} V_{\text{exptmax}}(\text{Succ}(s,a)) & \text{Player}(s) = \text{agent} \\ \displaystyle\sum_{a\in\text{Actions}(s)} \pi_{\text{opp}}(s,a) V_{\text{exptmax}}(\text{Succ}(s,a)) & \text{Player}(s) = \text{opp} \end{cases}$$

Remarque : expectimax est l'analogue de l'algorithme d'itération sur la valeur pour les MDPs.