1 Configuration triviale

1.1 Processus de décision markovien

Entrée : $S(\{i|i=0..N^2\}), \pi(S), \gamma, \epsilon$
Sortie : $\pi^\star(\mathbf{s}), \mathbf{v}^\star_\pi(S)$

```
# Initialisation
```
$\mathbf{v}^\star_{\pi,\mathbf{0}}(S) \leftarrow \mathbf{0}$
Pour $t = 1$ à t_{obs} **Faire**
  Pour s dans s **Faire**

$$\mathbf{v}_{\pi,\mathbf{t}}(S) \leftarrow \left(\sum_{s' \in S} p(s,a,s') \left(r(s,a,s') + \gamma v^\star_{\pi,t-1}(s') \right) \right)_{a \in \pi(S)}$$

$$\mathbf{v}^\star_{\pi,\mathbf{t}}(S) \leftarrow \max_{a \in \pi(S)} \mathbf{v}_{\pi,\mathbf{t}}(S)$$

$$\pi^\star(S) \leftarrow \left(\max_{a \in \pi(s)} \mathbf{v}_{\pi,\mathbf{t}}(s) \right)_{s \in S}$$

    Si $\max_{s \in S} \left(v^\star_{\pi,t}(s) - v^\star_{\pi,t-1}(s) \right) \leq \epsilon$ **Alors**
      Retourner $\pi^\star(S), \mathbf{v}^\star_{\pi,\mathbf{t}}(S)$
    Fin Si
  Fin Pour
Fin Pour
Retourner $\pi^\star(S), \mathbf{v}^\star_{\pi,\mathbf{t}}(S)$

1.2 SARSA

1.3 Q-learning

2 Intégration de pièges