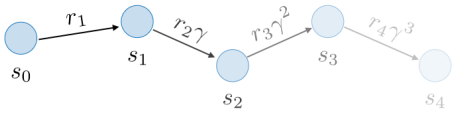
$$\forall s,a, \quad \boxed{\sum_{s' \in \text{ States}} T(s,a,s') = 1}$$

❒ **Politique** – Une politique π est une fonction liant chaque état s à une action a, i.e.

$$\boxed{\pi : s \mapsto a}$$

❒ **Utilité** – L'utilité d'un chemin $(s_0, ..., s_k)$ est la somme des récompenses dévaluées récoltées sur ce chemin. En d'autres termes,

$$\boxed{u(s_0,...,s_k) = \sum_{i=1}^{k} r_i \gamma^{i-1}}$$



Remarque : la figure ci-dessus illustre le cas $k = 4$.

❒ **Q-value** – La fonction de valeur des états-actions (*Q-value* en anglais) d'une politique π évaluée à l'état s avec l'action a, aussi notée $Q_\pi(s,a)$, est l'espérance de l'utilité partant de l'état s avec l'action a et adoptant ensuite la politique π. Cette fonction est définie par :

$$\boxed{Q_\pi(s,a) = \sum_{s' \in \text{ States}} T(s,a,s') \left[\text{Reward}(s,a,s') + \gamma V_\pi(s')\right]}$$

❒ **Fonction de valeur des états d'une politique** – La fonction de valeur des états d'une politique π évaluée à l'état s, aussi notée $V_\pi(s)$, est l'espérance de l'utilité partant de l'état s et adoptant ensuite la politique π. Cette fonction est définie par :

$$\boxed{V_\pi(s) = Q_\pi(s,\pi(s))}$$

Remarque : $V_\pi(s)$ vaut 0 si s est un état final.

2.2.2 Applications

❒ **Évaluation d'une politique** – Étant donnée une politique π, on peut utiliser l'algorithme itératif d'évaluation de politiques (en anglais *policy evaluation*) pour estimer V_π :

— Initialisation : pour tous les états s, on a

$$\boxed{V_\pi^{(0)}(s) \longleftarrow 0}$$

— Itération : pour t allant de 1 à T_{PE}, on a

$$\forall s, \quad \boxed{V_\pi^{(t)}(s) \longleftarrow Q_\pi^{(t-1)}(s,\pi(s))}$$

avec