— On itère la valeur en se basant sur les valeurs précédentes :

$$\boxed{V_{i+1}(s) = R(s) + \max_{a \in \mathcal{A}} \left[\sum_{s' \in \mathcal{S}} \gamma P_{sa}(s') V_i(s') \right]}$$

❒ **Maximum de vraisemblance** – Les estimations du maximum de vraisemblance pour les transitions de probabilité d'état sont comme suit :

$$\boxed{P_{sa}(s') = \frac{\text{\#fois où l'action } a \text{ dans l'état s } s \text{ est prise pour arriver à l'état } s'}{\text{\#fois où l'action } a \text{ dans l'état } s \text{ est prise}}}$$

❒ **Q-learning** – Le Q-learning est une estimation non-paramétrique de Q, qui est faite de la manière suivante :

$$\boxed{Q(s,a) \leftarrow Q(s,a) + \alpha \left[R(s,a,s') + \gamma \max_{a'} Q(s',a') - Q(s,a) \right]}$$