

Bootstrapping en Apprentissage par Renforcement

En RL, **bootstrapping** veut dire :

Mettre à jour une estimation à partir d'une autre estimation, pas uniquement à partir de « vraies » données complètes.

1. Dans les équations TD

Exemple TD(1) :

$$V(S_t) \leftarrow (1 - \alpha)V(S_t) + \alpha [R_{t+1} + \gamma V(S_{t+1})] \quad (1)$$

— Partie **récompense réelle** : R_{t+1}

— Partie **bootstrapping** : $\gamma V(S_{t+1})$

Tu n'attends pas la fin de l'épisode pour connaître le « vrai » retour total. Tu dis :

« je vais utiliser ma meilleure estimation actuelle de $V(S_{t+1})$ pour corriger $V(S_t)$ »

Donc :

bootstrapping = utiliser $V(S_{t+1})$, $V(S_{t+2})$, etc., qui sont eux-mêmes approxi-
més, comme cibles partielles.

2. Différence avec Monte Carlo

Monte Carlo : pas de bootstrapping

$$V(S_t) \leftarrow V(S_t) + \alpha(G_t - V(S_t)) \quad (2)$$

où G_t = retour complet jusqu'à la fin de l'épisode (somme de vraies récompenses).

TD / Q-Learning : bootstrapping

Les cibles sont du type :

$$R_{t+1} + \gamma V(S_{t+1}) \quad \text{ou} \quad R_{t+1} + \gamma \max_a Q(S_{t+1}, a) \quad (3)$$

Donc tu mets à jour **en t'appuyant sur tes propres estimations**.

Résumé

- **Bootstrapping** = corriger une valeur avec une cible qui contient déjà des valeurs estimées.
- C'est plus **rapide** (pas besoin d'attendre la fin) mais plus **biaisé** que Monte Carlo.