Model-based (tabular) RL:

For each (s,a), we have $\{(r_i, s_i')\}_{i=1}^n$ — $r_i \sim R(s,a)$, $s_i' \sim P(\cdot|s,a)$.

Build empirical model $\hat{M}$.

$$\hat{R}(s,a) = \frac{1}{n}\sum_{i=1}^{n} r_i \approx R(s,a)$$

$$\hat{P}(s'|s,a) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[s_i' = s'] \approx P(\cdot|s,a)$$

Run Value-iteration in $\hat{M}$.

$$(\hat{\mathcal{T}} f)(s,a) = \hat{R}(s,a) + \gamma \mathbb{E}_{s' \sim \hat{P}(\cdot|s,a)}\left[\max_{a'} f(s',a')\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} r_i + \gamma \sum_{s' \in S} \hat{P}(s'|s,a) \cdot \left(\max_{a'} f(s',a')\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} r_i + \gamma \sum_{s' \in S} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[s_i' = s']\right) \cdot \max_{a'} f(s',a')$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} r_i + \gamma \sum_{s' \in S} \sum_{i=1}^{n} \mathbb{I}[s_i' = s'] \max_{a'} f(s',a')\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} r_i + \gamma \sum_{i=1}^{n} \max_{a'} f(s_i',a')\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(r_i + \gamma \max_{a'} f(s_i',a')\right) \approx (\mathcal{T} f)(s,a)$$

empirical Bellman update.

For fixed f.

$$\mathbb{E}_{\substack{r_i \sim R(\cdot|s,a) \\ s_i' \sim P(\cdot|s,a)}}\left[r_i + \gamma \max_{a'} f(s_i', a')\right] = R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}\left[\max_{a'} f(s',a')\right] = (\mathcal{T} f)(s,a).$$