

Algorithm 1 Categorical Algorithm

input A transition $x_t, a_t, r_t, x_{t+1}, \gamma_t \in [0, 1]$

$$Q(x_{t+1}, a) := \sum_i z_i p_i(x_{t+1}, a)$$

$$a^* \leftarrow \arg \max_a Q(x_{t+1}, a)$$

$$m_i = 0, \quad i \in 0, \dots, N-1$$

for $j \in 0, \dots, N-1$ **do**

Compute the projection of $\hat{\mathcal{T}} z_j$ onto the support $\{z_i\}$

$$\hat{\mathcal{T}} z_j \leftarrow [r_t + \gamma_t z_j]_{V_{\min}}^{V_{\max}}$$

$$b_j \leftarrow (\hat{\mathcal{T}} z_j - V_{\min}) / \Delta z \quad \# b_j \in [0, N-1]$$

$$l \leftarrow \lfloor b_j \rfloor, u \leftarrow \lceil b_j \rceil$$

Distribute probability of $\hat{\mathcal{T}} z_j$

$$m_l \leftarrow m_l + p_j(x_{t+1}, a^*)(u - b_j)$$

$$m_u \leftarrow m_u + p_j(x_{t+1}, a^*)(b_j - l)$$

end for

output $-\sum_i m_i \log p_i(x_t, a_t)$ # Cross-entropy loss