- Với một trạng thái s, sai số xấp xỉ là:

$$\text{Error}(s) = (V^\pi(s) - \hat{v}(s; \mathbf{w}))^2$$

→ gọi là **squared error**.

Nhưng: chỉ lỗi tại một trạng thái là chưa đủ. Cần đo **tổng thể trên toàn bộ không gian trạng thái**.

3. Mean Squared Value Error (MSVE)

MSVE là **trung bình có trọng số** của squared error trên mọi trạng thái:

$$\text{MSVE}(\mathbf{w}) = \sum_{s \in \mathcal{S}} \mu(s) \left(V^\pi(s) - \hat{v}(s; \mathbf{w})\right)^2$$

Trong đó:

- $\mu(s)$: **phân phối trạng thái** – đại diện cho tầm quan trọng của từng trạng thái.
- $\hat{v}(s; \mathbf{w})$: giá trị xấp xỉ từ hàm học.
- $V^\pi(s)$: giá trị thực tế của trạng thái dưới chính sách π.

Từ Monte Carlo đến TD Learning

- **Monte Carlo update**:

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(G_t - \hat{v}(S_t; \mathbf{w}))\nabla_{\mathbf{w}} \hat{v}(S_t; \mathbf{w})$$

  - G_t: Return đầy đủ (tổng phần thưởng đến cuối episode)
- Nhưng bạn **không cần chờ đến cuối episode** → dùng **TD target**:

$$U_t = R_{t+1} + \gamma \hat{v}(S_{t+1}; \mathbf{w})$$

  - Đây là **bootstrap estimate** (ước lượng từ chính hàm giá trị hiện tại)