Sparse Tree Search Optimality Guarantees in POMDPs with Continuous Observation Spaces

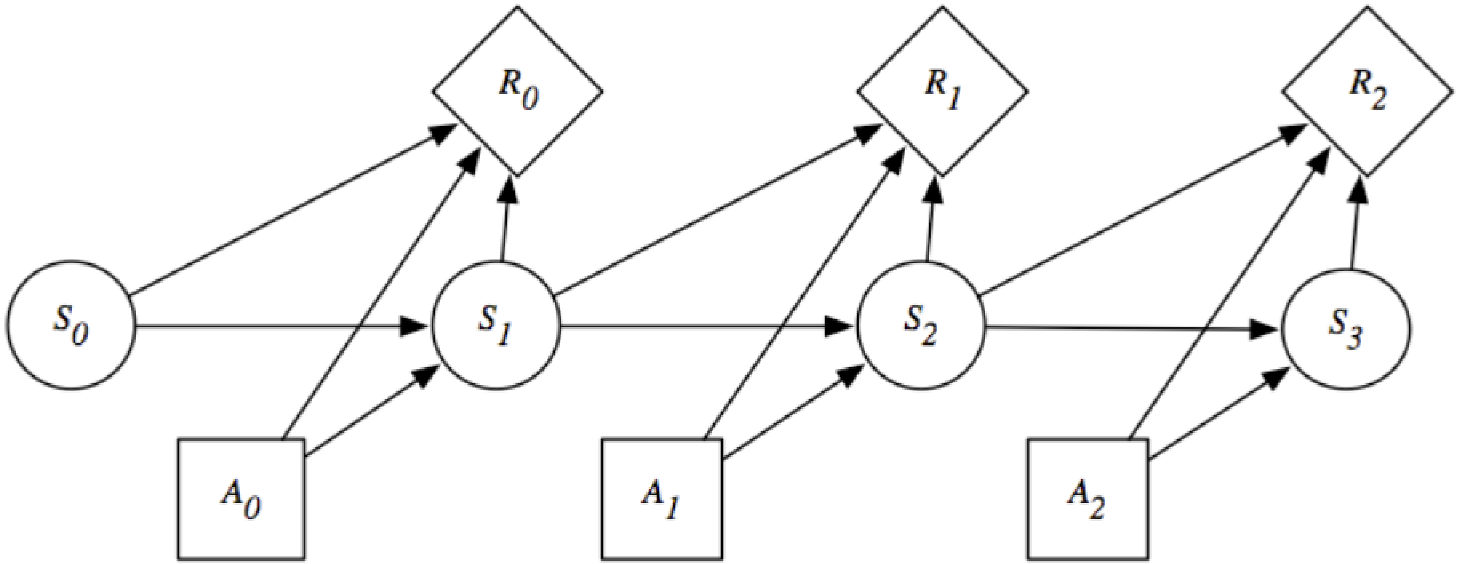
Tianci Li

2021/8/27

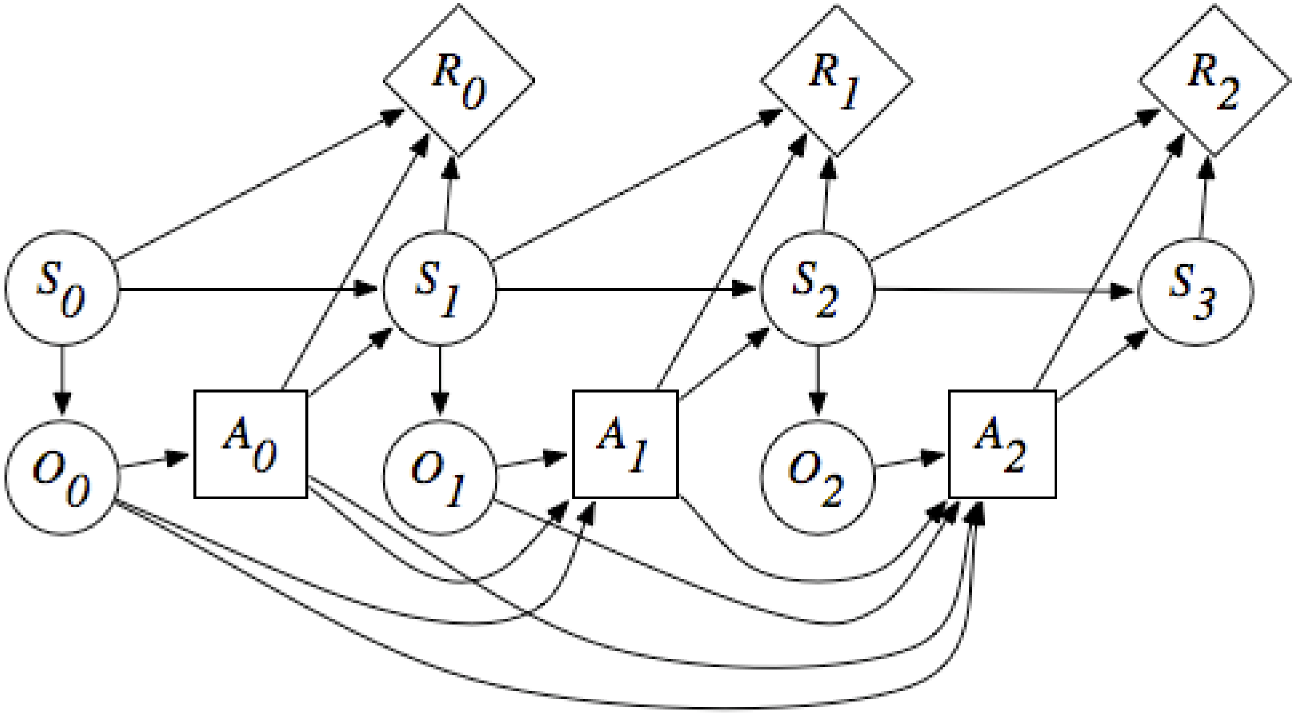
Outline

- Introduction
  - Definition of POMDP
  - Sparse Tree Search Algorithms in POMDP
    - Monte-Carlo Tree Search
    - Particle Filtering
    - Progressive widening
    - Importance Sampling
- Sparse Tree Search Optimality Guarantees

Introduction



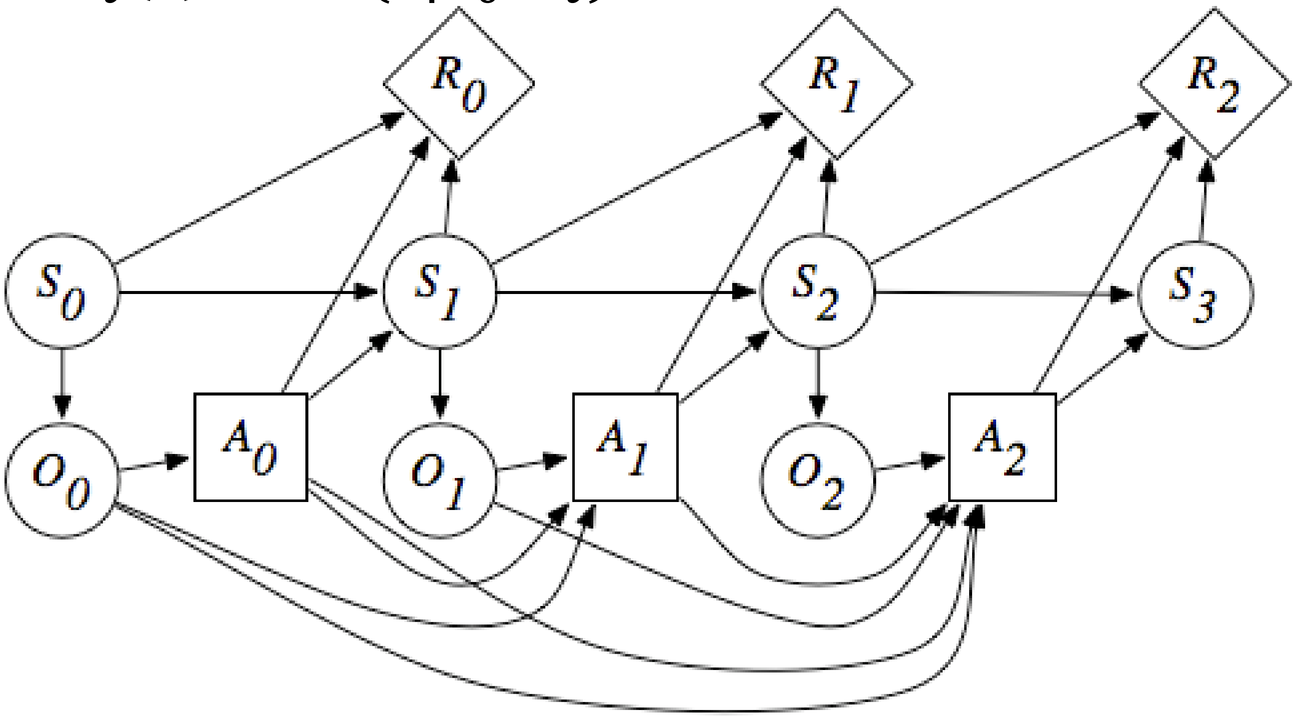
Introduction



Introduction

历史信息：$h_t = \{a_0, o_1, a_2, \ldots, a_{t-1}, o_t\}$

信念状态：$b_t(s) = Pr(s|b_0, h_t)$



POMDP → Belief MDP

Definition of POMDP

A POMDP can formally be described as a 7-tuple $\mathrm{P} = (\mathcal{S}, \mathcal{A}, \mathrm{T}, \mathrm{R}, \Omega, \mathrm{O}, \gamma)$,

- $\mathcal{S} = \{\mathrm{s}_1, \mathrm{s}_2, \dots, \mathrm{s}_\mathrm{n}\}$ is a set of states,
- $\mathcal{A} = \{\mathrm{a}_1, \mathrm{a}_2, \dots, \mathrm{a}_\mathrm{m}\}$ is a set of actions,
- T is a set of conditional transition probabilities $T(s' \mid s, a)$ for the state transition $s \rightarrow s'$.
- $R: S \times A \rightarrow R$ is the reward function,
- $\Omega = \{\mathrm{o}_1, \mathrm{o}_2, \dots, \mathrm{o}_\mathrm{k}\}$ is a set of observations,
- O is a set of observation probabilities $\mathrm{O}(\mathrm{o}|\mathrm{s}', \mathrm{a})$
- $\gamma \in [0,1]$ is the discount factor.

Some concepts and equations

历史信息：$h_t = \{a_0, o_1, a_2, \dots, a_{t-1}, o_t\}$

信念状态：$b_t(s) = Pr(s|b_0, h_t)$

多变量贝叶斯公式：

$$\Pr(A|B,C) = \frac{\Pr(B|A,C)\Pr(A|C)}{\Pr(B|C)}$$

贝叶斯更新：$b_t(s') = Pr(s'|a_{t-1}, o_t, b_{t-1})$

$$= \frac{\Pr(o_t|s', a_{t-1}, b_{t-1})\Pr(s'|a_{t-1}, b_{t-1})}{\Pr(o_t|a_{t-1}, b_{t-1})}$$

$$= \frac{\mathrm{O}(o_t|s', a_{t-1})\sum_{s\in S} T(s'|s, a_{t-1}) b_{t-1}(s)}{\sum_{s'\in S} O(o_t|s', a_{t-1})\sum_{s\in S} T(s'|s, a_{t-1}) b_{t-1}(s)}$$

$O(|S|^2)$

计算成本高！使用粒子滤波

Some concepts and equations

累计奖励： $G_t = R_t + \gamma R_{t+1} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$

策略：$\forall s \in S, \forall a \in A, \pi(a|s) = \Pr(a|s)$

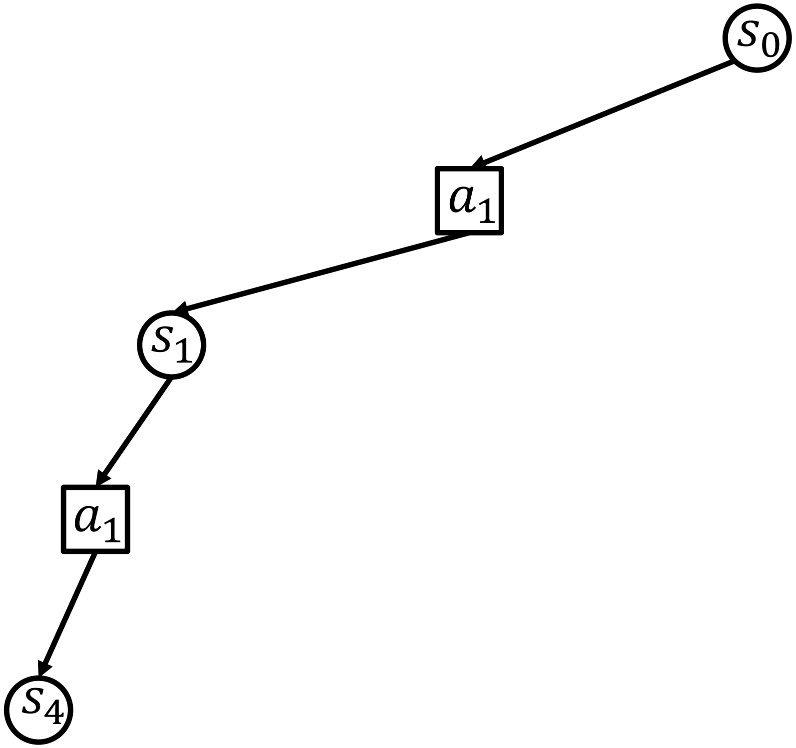
状态值函数：$V_\pi(s) = E[G_t|s_t = s] = E[R_t + \gamma V_\pi(s_{t+1})|s_t = s]$(贝尔曼方程)

动作值函数：$Q_\pi(s,a) = R(s,a) + \gamma \sum_{s' \in S} P(s'|s,a) V_\pi(s')$

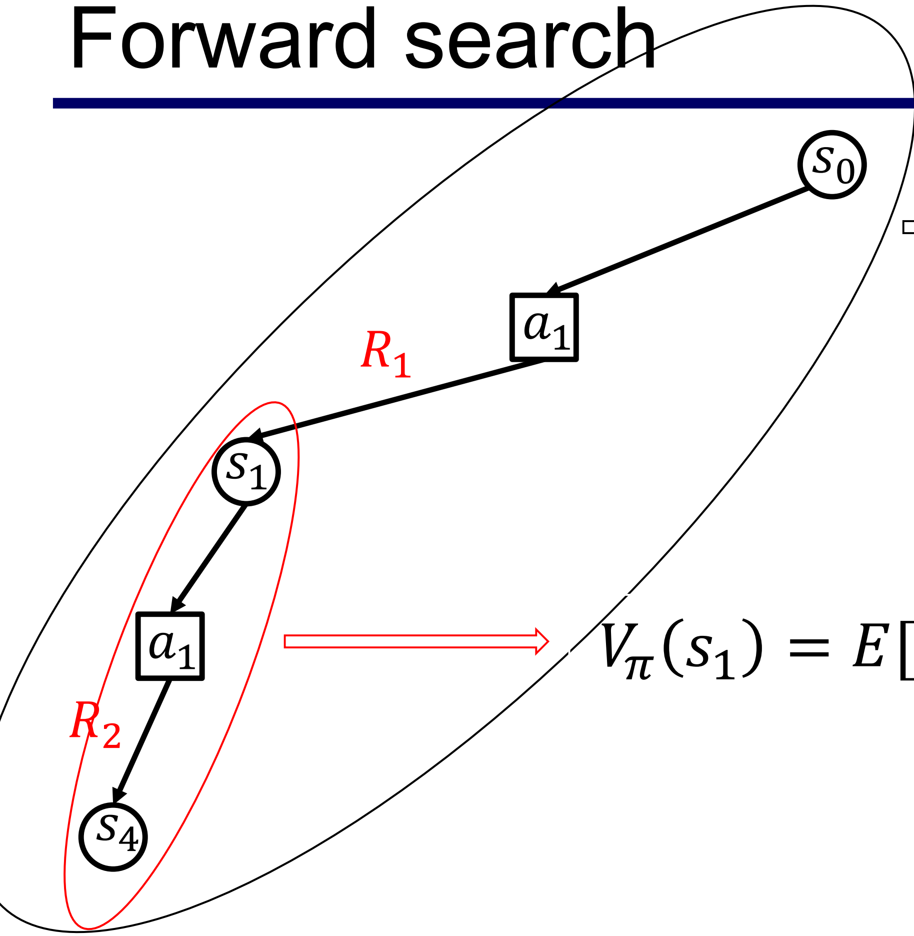
贝尔曼最优方程：$V^*(s) = \max_a Q(s,a)$

最优策略：$\pi^*(a|s) = \begin{cases} 1, & if\ a \in arg\max_a Q(s,a) \\ 0, & otherwise \end{cases}$

Forward search



Forward search

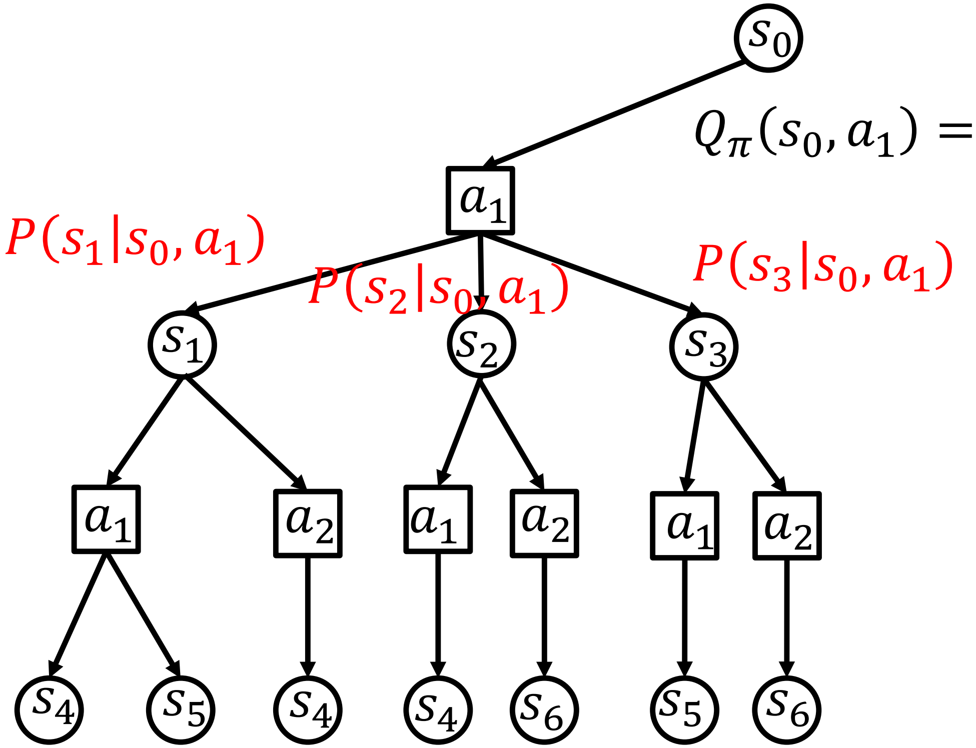


$$V_\pi(s_0) = E[G_0]$$
$$= E[R_1 + \gamma G_1]$$
$$= E[R_1 + \gamma V_\pi(s_1)]$$

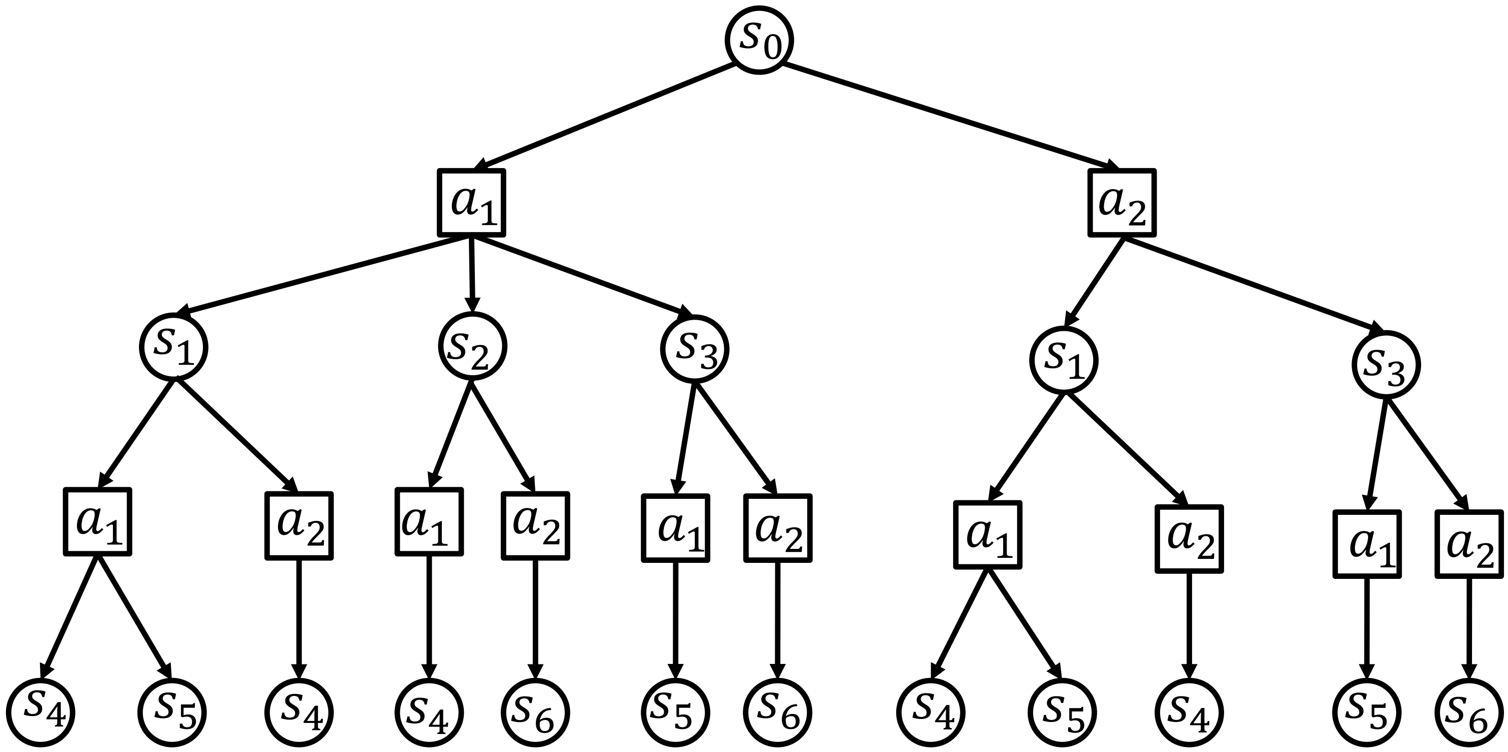
$$V_\pi(s_1) = E[G_1]$$

Forward search

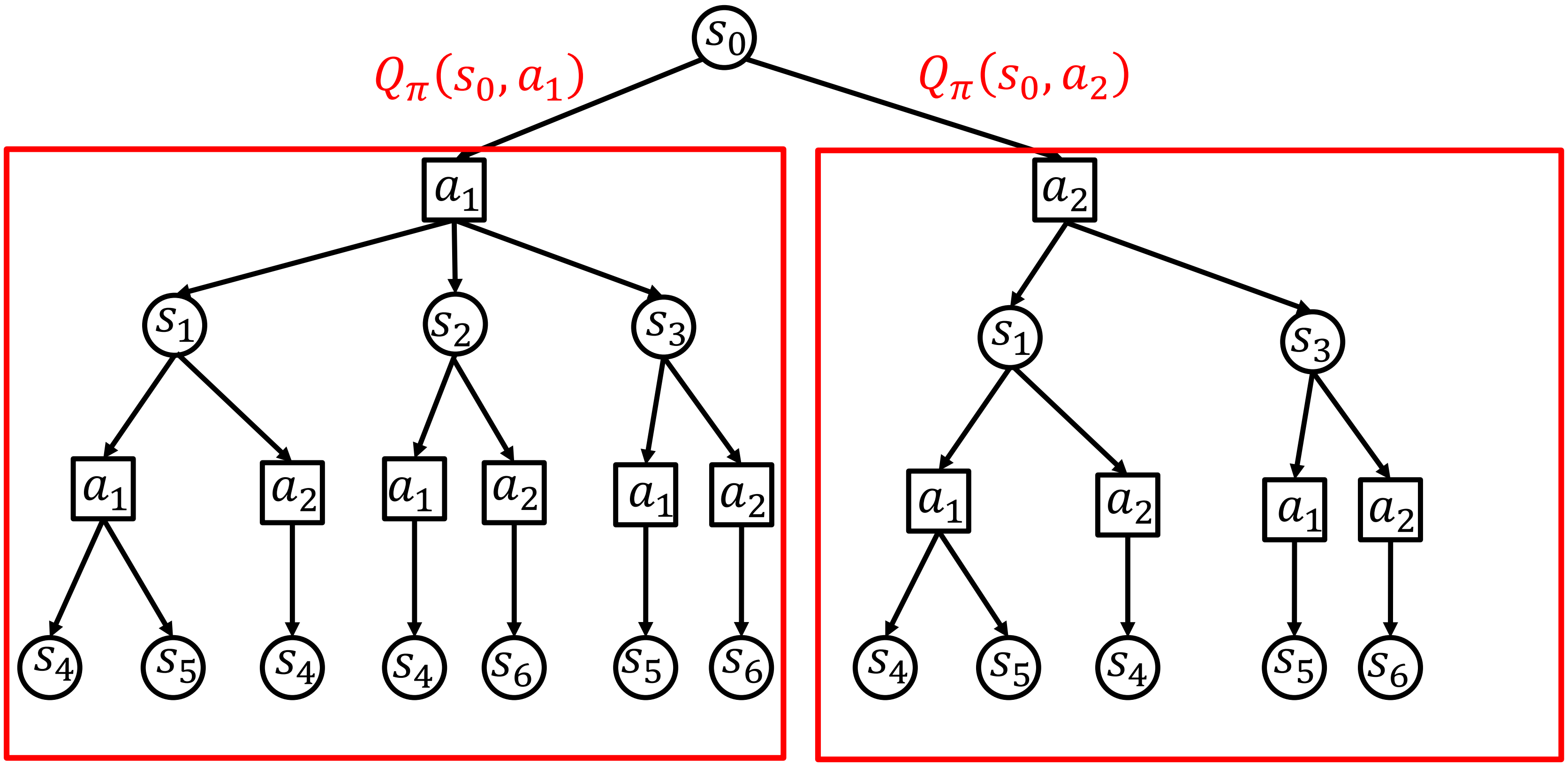
$$Q_\pi(s_0, a_1) = R(s_0, a_1) + \gamma \sum_{s' \in S} P(s'|s_0, a_1) V_\pi(s')$$



Forward search

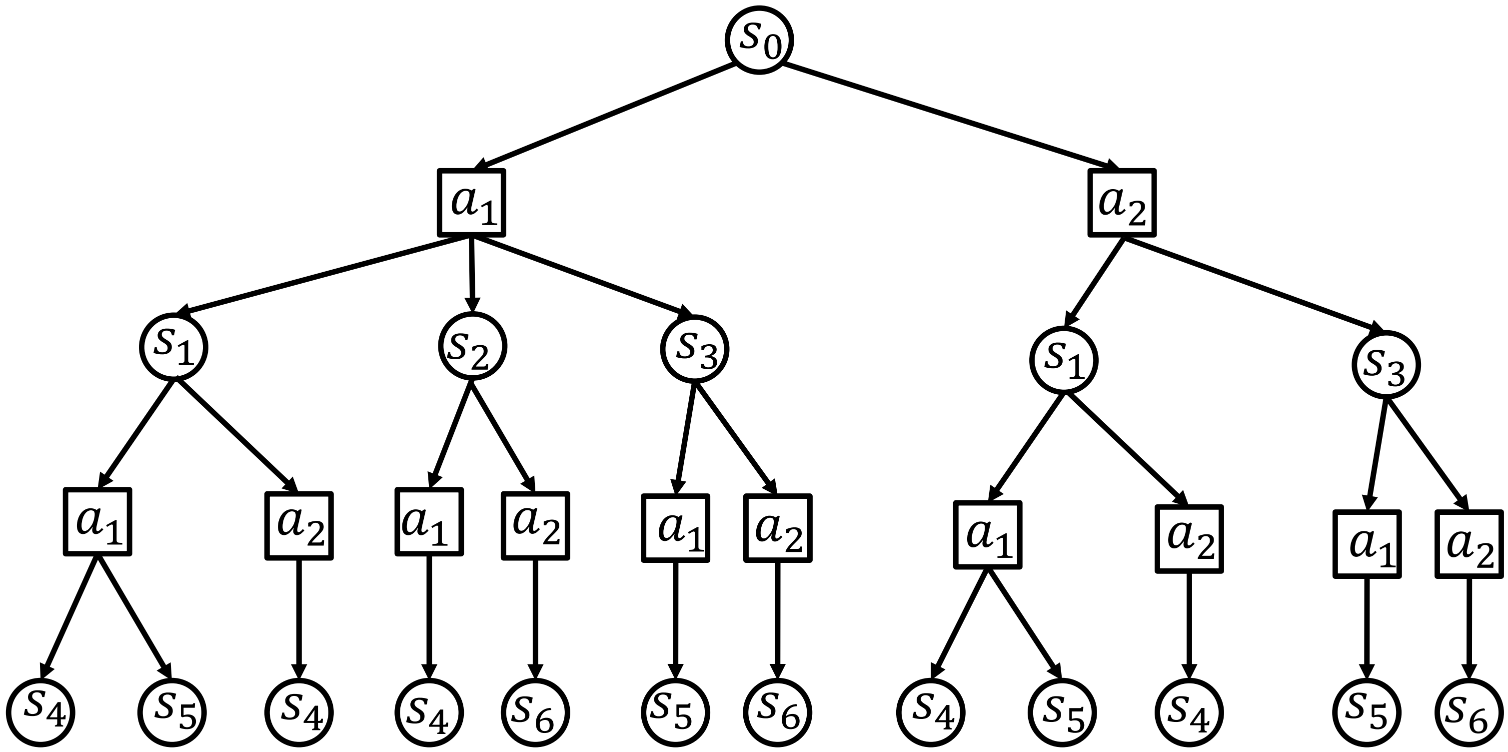


Forward search

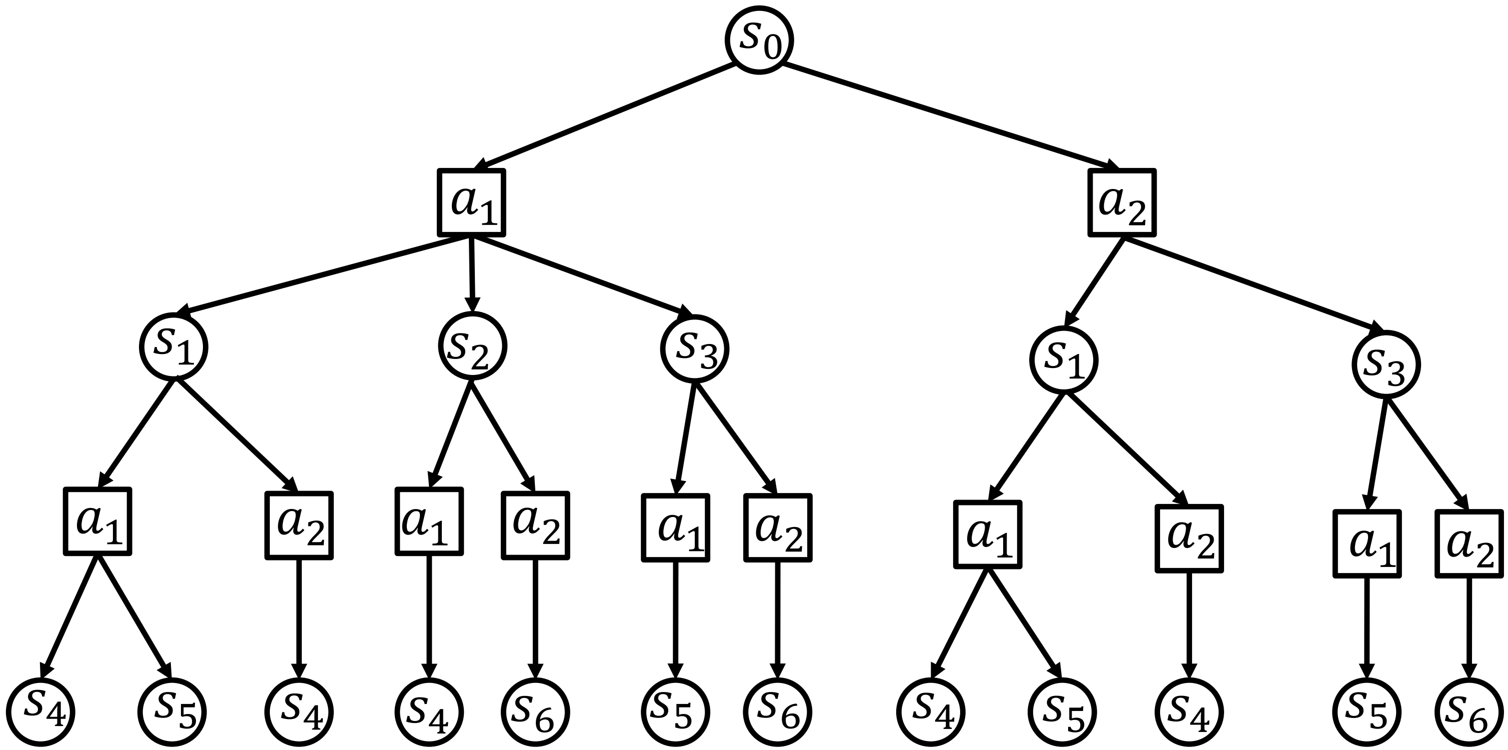


$$V_\pi(s_0) = \sum_{a \in A} \pi(a|s_0)\, Q_\pi(s_0, a) \le \max_a Q(s_0, a) \qquad V^*(s_0) = \max_a Q^*(s_0, a)$$

Forward search



Forward search

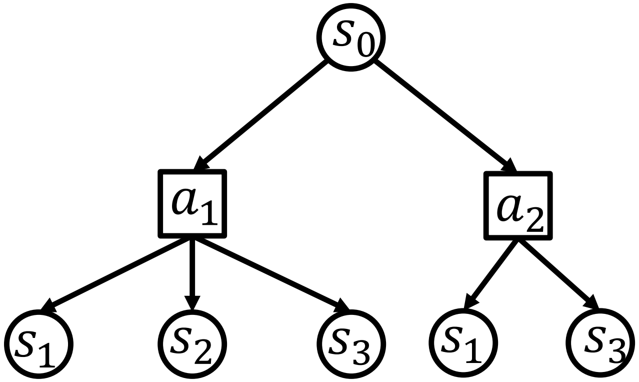


When the search space is large

Forward search ✗ Monte-Carlo Tree Search ✓

Monte-Carlo Tree Search

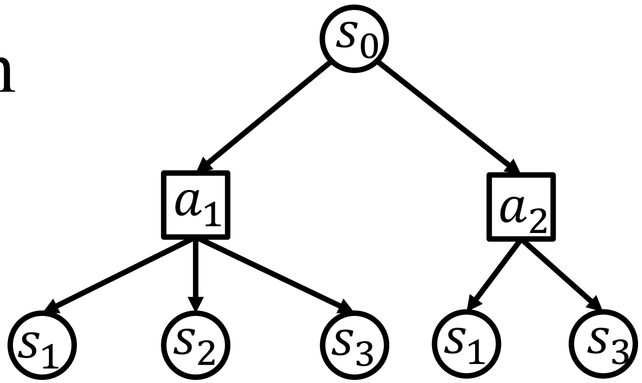
Each node saves the N value(visited counts) and V value(Node value)



Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)

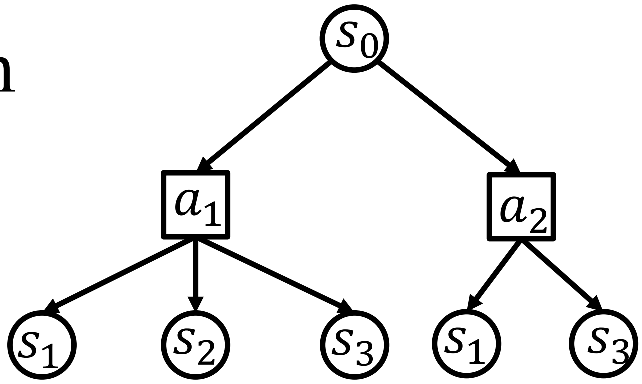
1.Selection



Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)

1.Selection



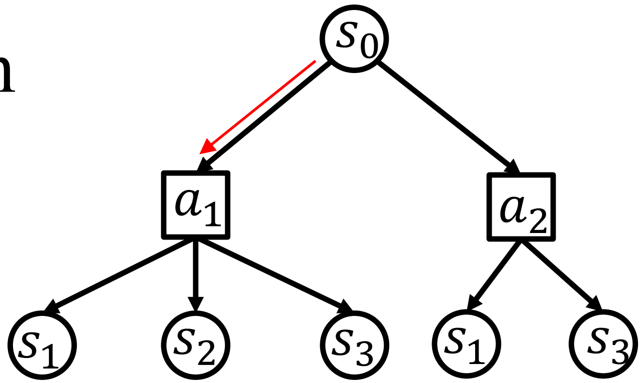
A tree policy that is used while within the search tree. Greedy tree policy ; UCT

UCT: $$A = \underset{a}{argmax}\ [V(sa) + c\sqrt{\frac{\log N(s)}{N(sa)}}\ ]$$

Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)

1.Selection

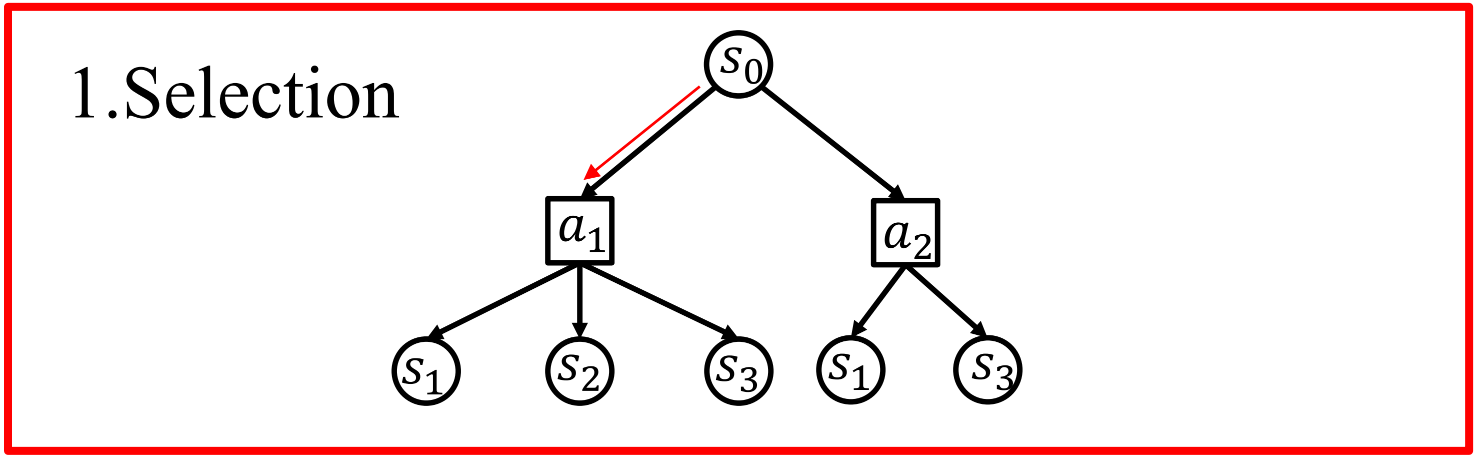


A tree policy that is used while within the search tree. Greedy tree policy ; UCT

$$\text{UCT:} \quad A = \underset{a}{\arg\max} \left[V(sa) + c\sqrt{\frac{\log N(s)}{N(sa)}} \right]$$

Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)



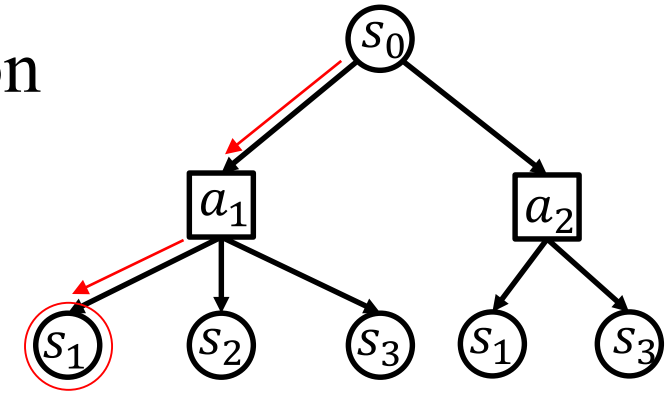
Generative model to generate sequences of states and rewards

$$G(s_t, a_t) = (R_{t+1}, s_{t+1})$$

Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)

1.Selection

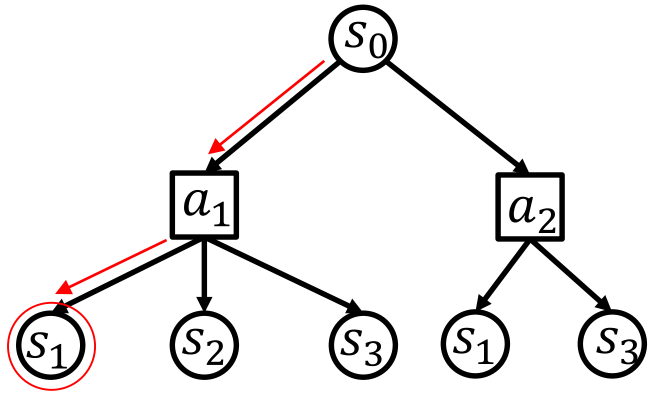


Generative model to generate sequences of states and rewards

$$G(s_t, a_t) = (R_{t+1}, s_{t+1})$$

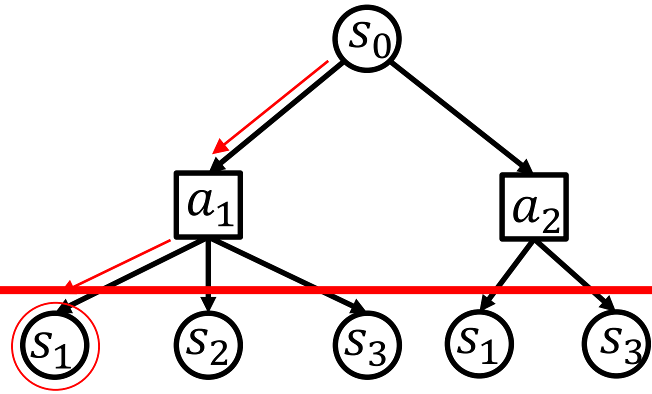
Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)



Monte-Carlo Tree Search

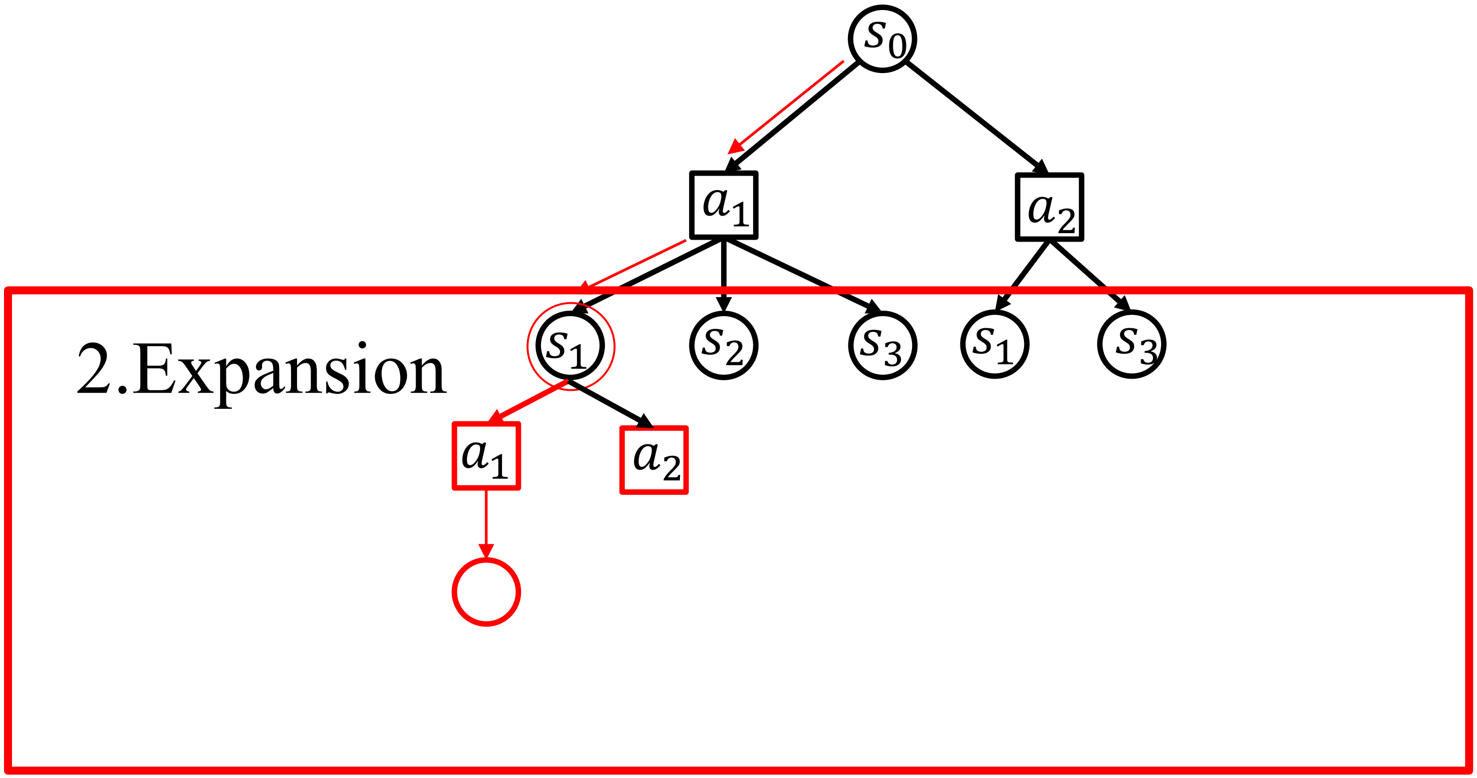
Each node saves the N value(visited counts) and V value(Node value)



2.Expansion

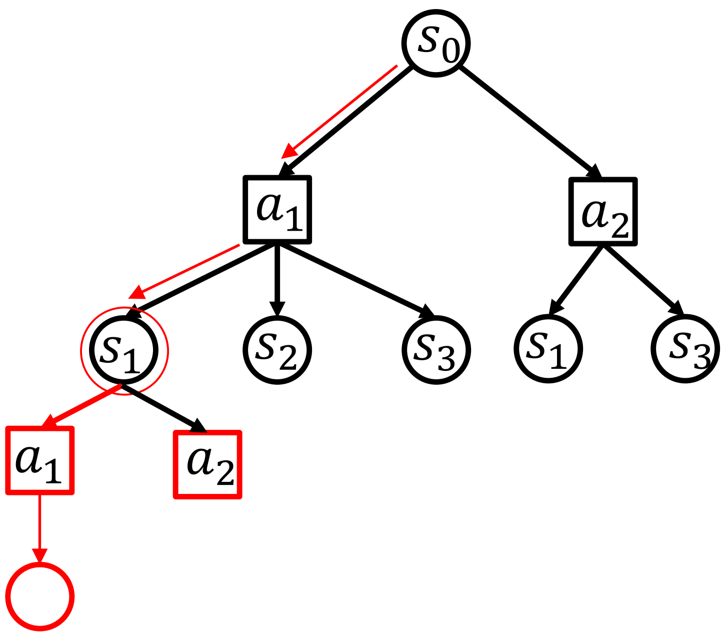
Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)



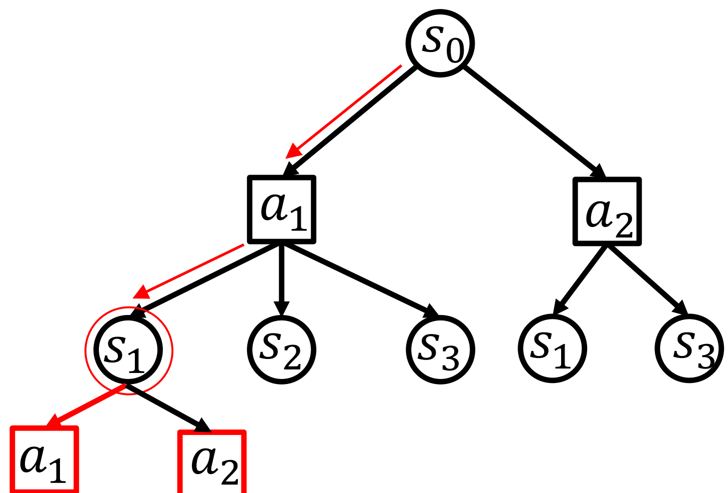
Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)

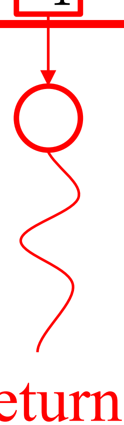


Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)



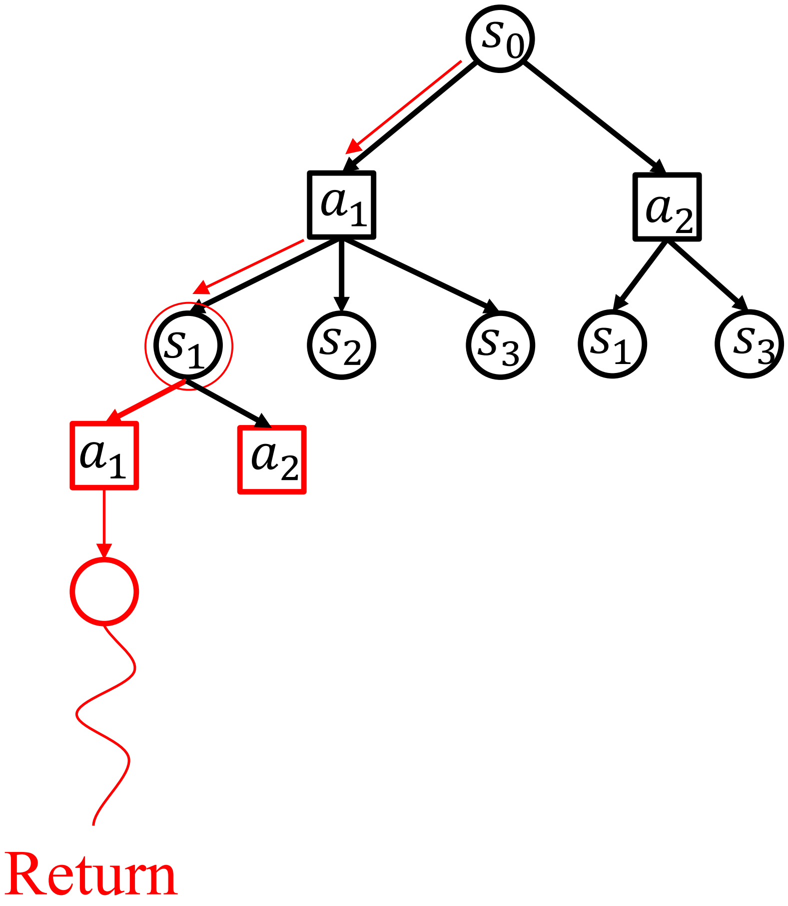
3.Simulation



A rollout policy that is used once simulations leave the scope of the search tree

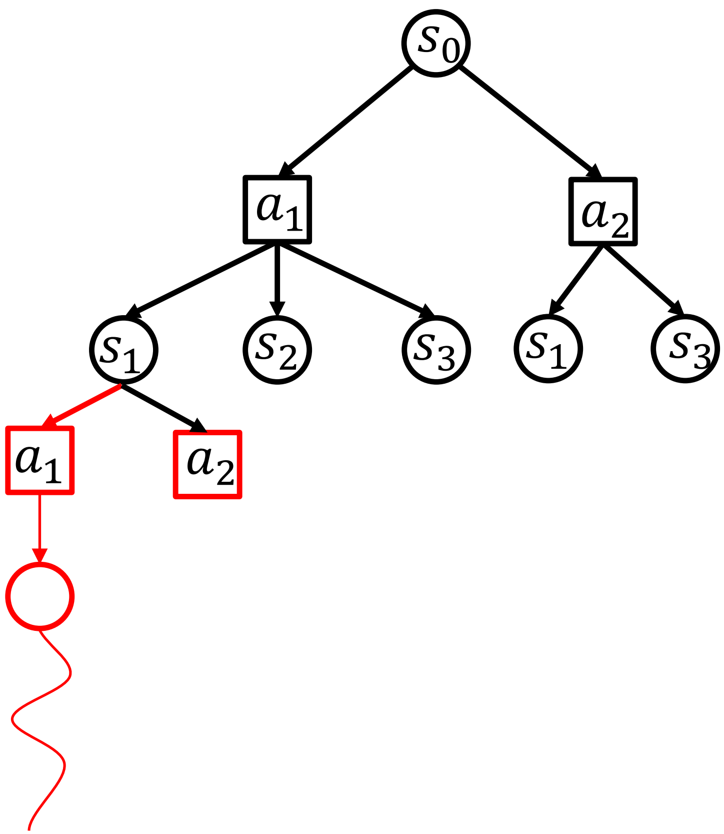
Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)



Monte-Carlo Tree Search

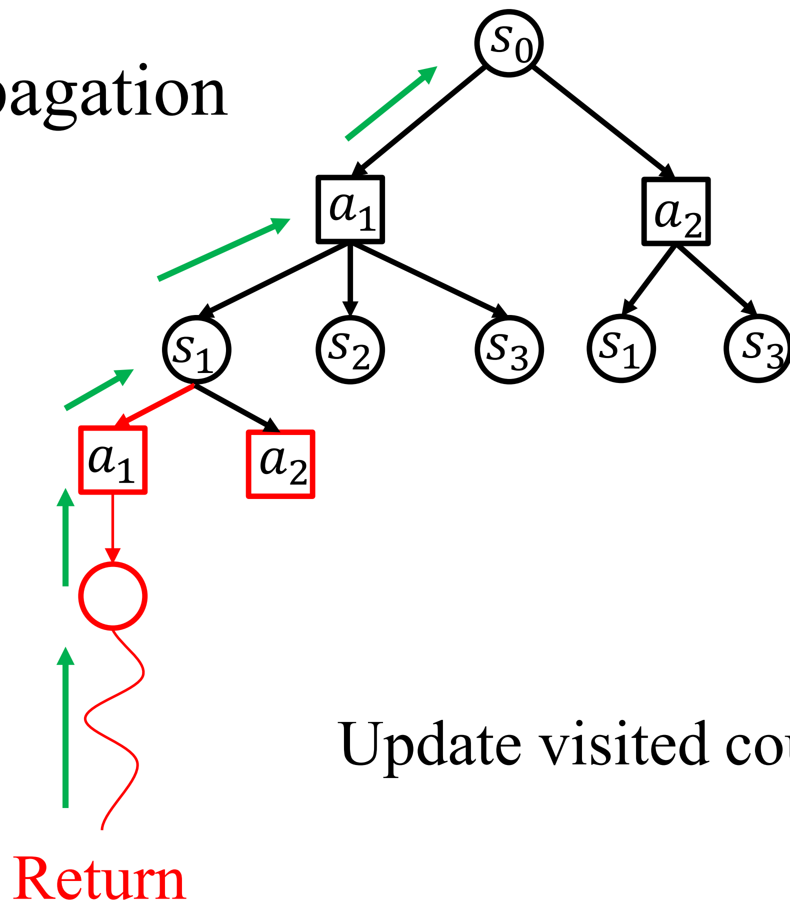
Each node saves the N value(visited counts) and V value(Node value)



Monte-Carlo Tree Search

Each node saves the N value(visited counts) and V value(Node value)

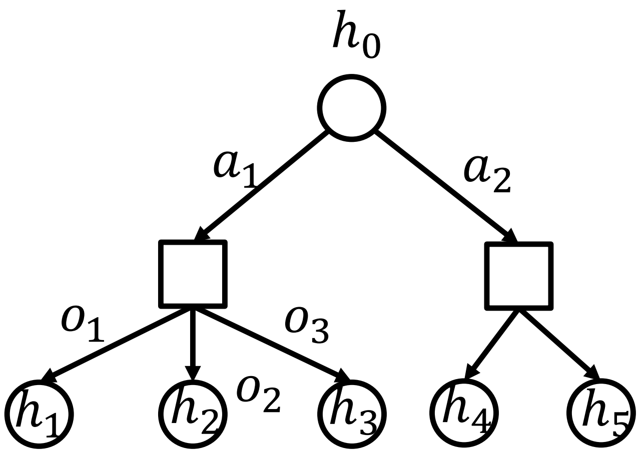
4.Backpropagation



Update visited counts and node value

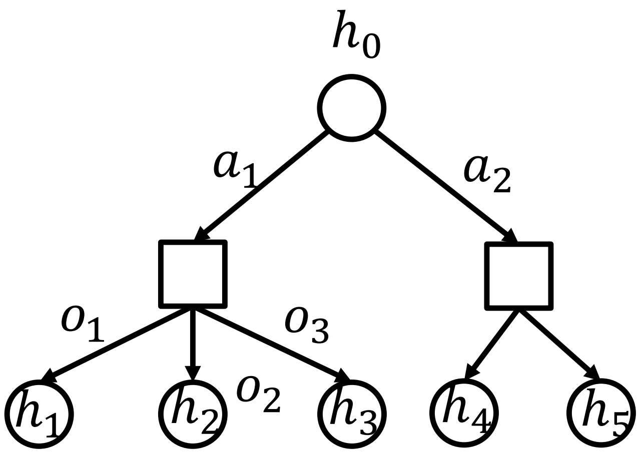
Monte-Carlo Tree Search

In POMCP:



Monte-Carlo Tree Search

In POMCP:

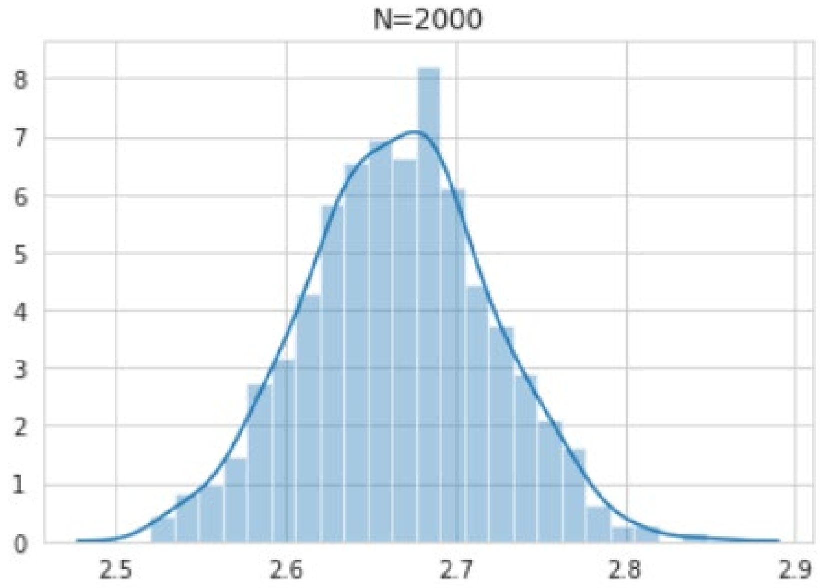


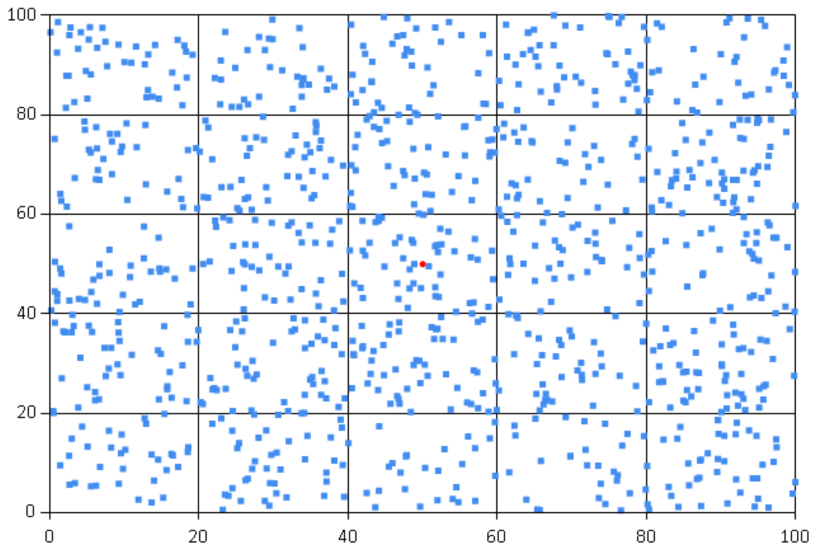
PO-UCT: $$A = \underset{a}{argmax}\ [V(ha) + c\sqrt{\frac{\log N(h)}{N(ha)}}]$$

Generative model: $$G(s_t, a_t) = (R_{t+1}, o_{t+1}, s_{t+1})$$

Particle Filtering

Using a set of particles to represent the posterior distribution.

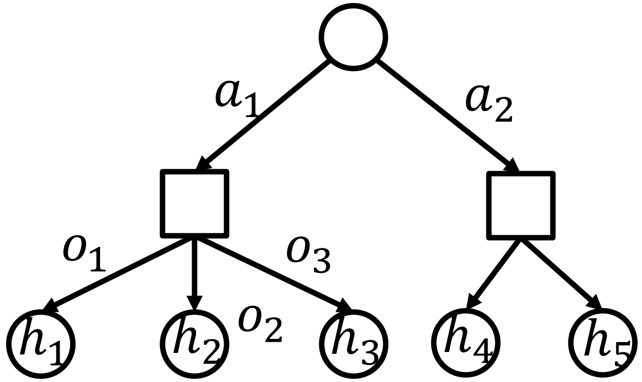




Particle Filtering

In POMCP:

h_0 $\mathrm{B}(h_0) = [s_0, s_0, s_1, s_2]$

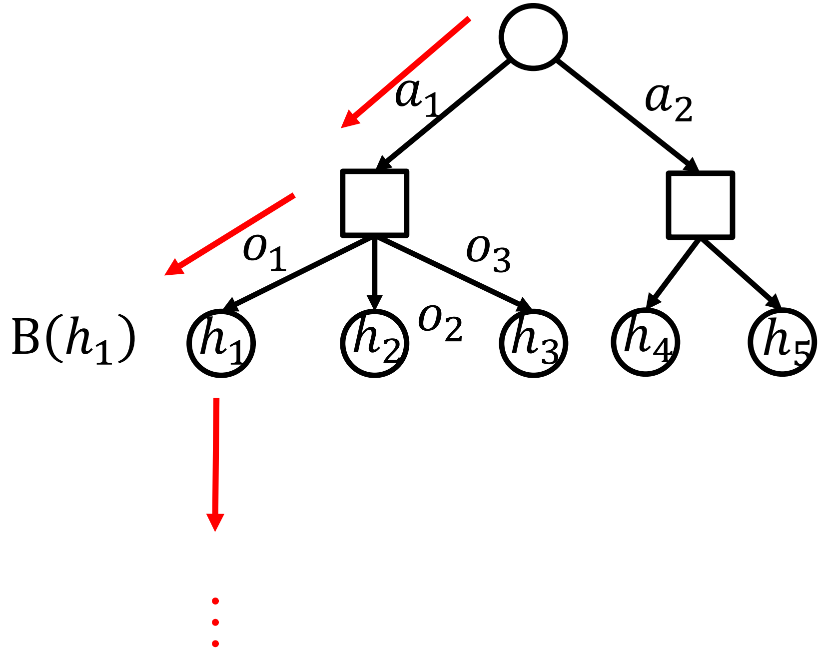


$\mathrm{B}(h_1)$

Particle Filtering

In POMCP:

h_0 $\mathrm{B}(h_0) = [s_0, s_0, s_1, s_2]$



$s \sim B(h_0)$

↓

$G(s, a_0) = (R_1, o_1, s_1)$

↓

$B(h_0 a_0 o_1) \cup \{s_1\}$

↓

$s \sim B(h_0 a_0 o_1)$

⋮

Partially Observable Monte-Carlo

Algorithm 1 Partially Observable Monte-Carlo Planning

```
procedure SEARCH(h)
    repeat
        if h = empty then
            s ~ I
        else
            s ~ B(h)
        end if
        SIMULATE(s, h, 0)
    until TIMEOUT()
    return argmax_b V(hb)
end procedure

procedure ROLLOUT(s, h, depth)
    if γ^depth < ε then
        return 0
    end if
    a ~ π_rollout(h, ·)
    (s', o, r) ~ G(s, a)
    return r + γ.ROLLOUT(s', hao, depth+1)
end procedure

procedure SIMULATE(s, h, depth)
    if γ^depth < ε then
        return 0
    end if
    if h ∉ T then
        for all a ∈ A do
            T(ha) ← (N_init(ha), V_init(ha), ∅)
        end for
        return ROLLOUT(s, h, depth)
    end if
    a ← argmax_b V(hb) + c sqrt(log N(h) / N(hb))
    (s', o, r) ~ G(s, a)
    R ← r + γ.SIMULATE(s', hao, depth + 1)
    B(h) ← B(h) ∪ {s}
    N(h) ← N(h) + 1
    N(ha) ← N(ha) + 1
    V(ha) ← V(ha) + (R − V(ha)) / N(ha)
    return R
end procedure
```

POMCP in Continues Space

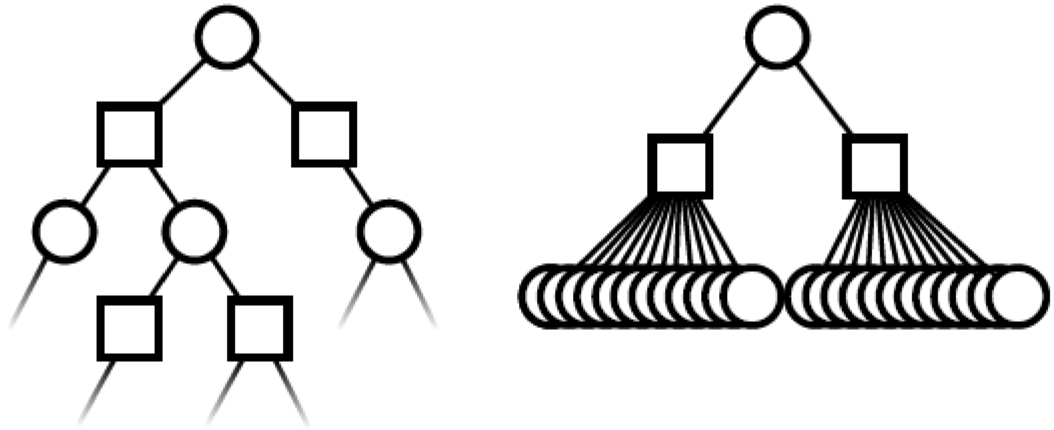
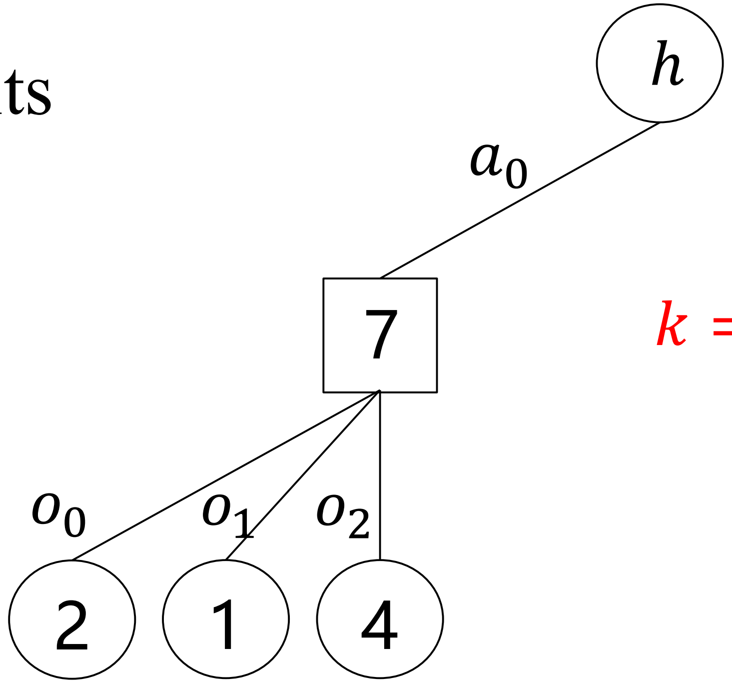


Figure 1: POMCP tree for a discrete POMDP (left), and for a POMDP with a continuous observation space (right). Because the observation space is continuous, each simulation creates a new observation node and the tree cannot extend deeper.

Progressive widening

- Limit the number of child nodes: $\leq kN^{\alpha}$

  N is the number of visits

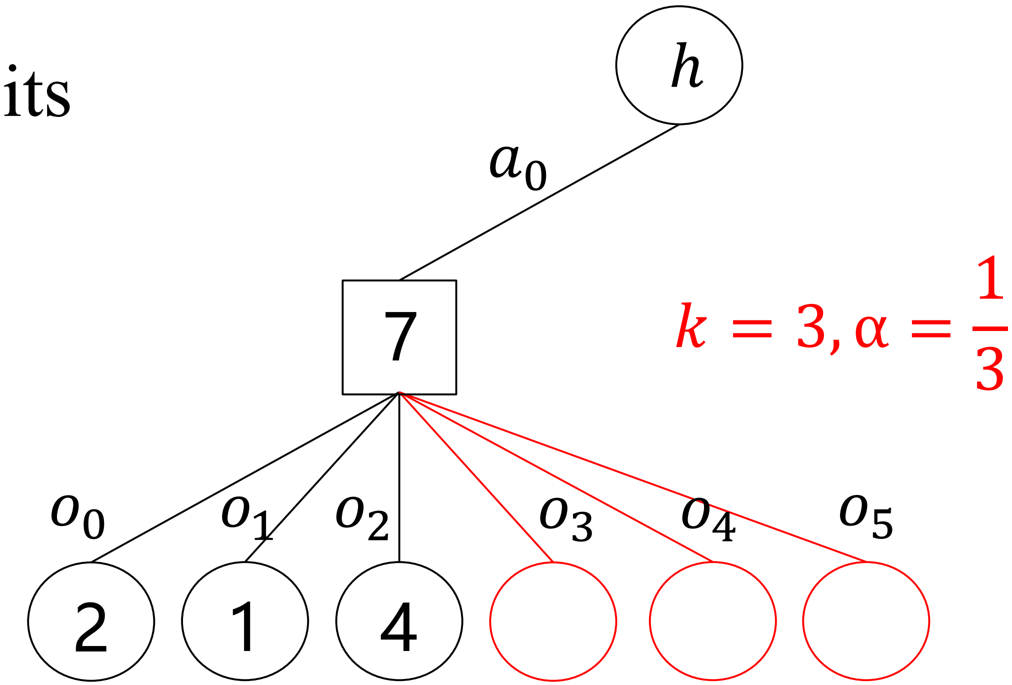


$k = 3, \alpha = \frac{1}{3}$

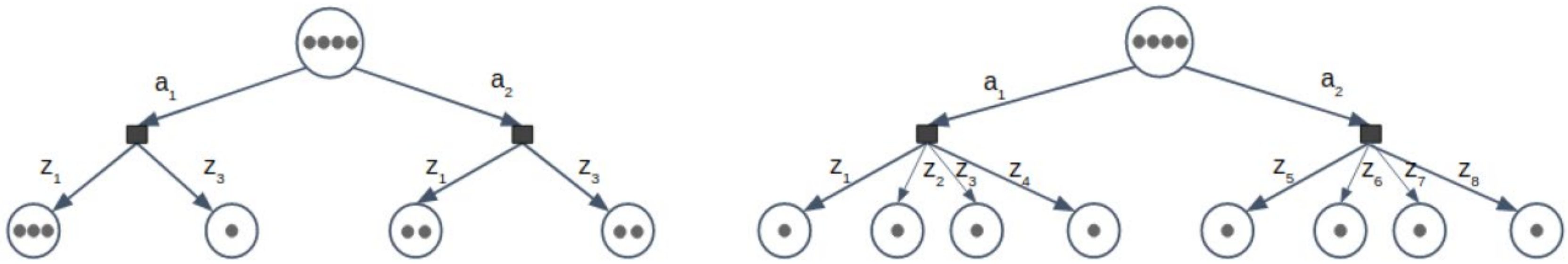
Progressive widening

- Limit the number of child nodes: $\leq kN^{\alpha}$

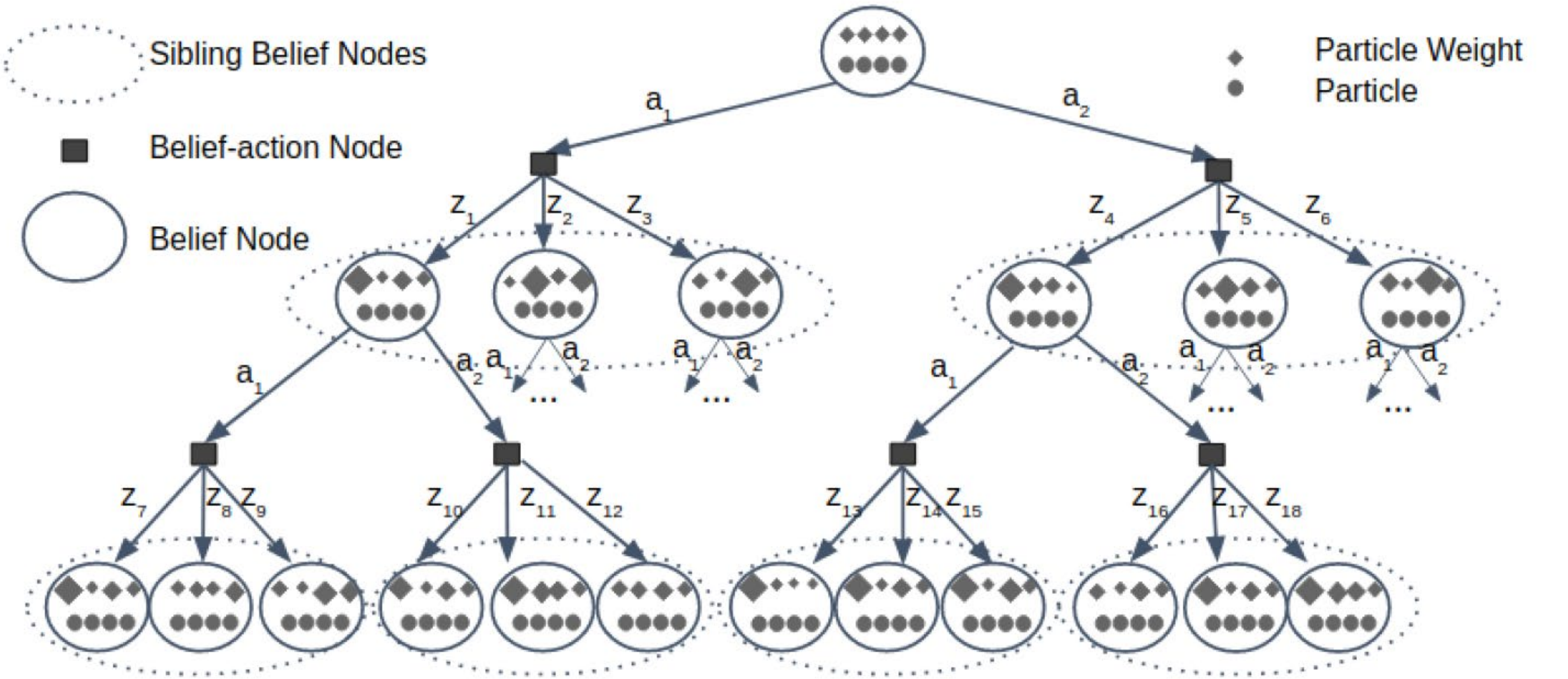
  N is the number of visits



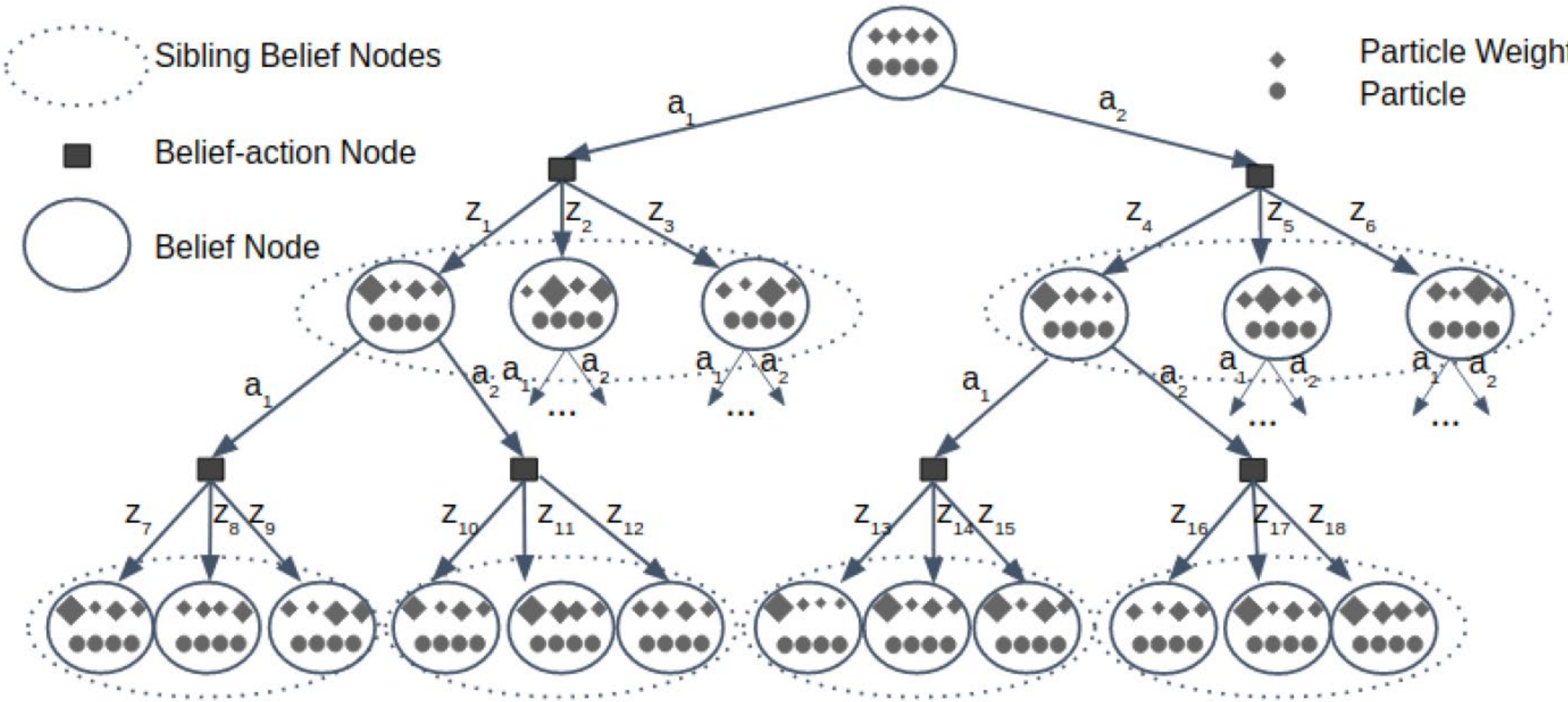
Belief Update



Belief Update



Belief Update



State particles are shared among all observation branches and importance sampling weights are given

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

$$E_p[f(x)] = \int f(x)p(x)dx = \int \frac{f(x)p(x)}{g(x)} g(x)dx = \frac{E_g\left[\frac{f(x)p(x)}{g(x)}\right]}{E_g[\frac{p(x)}{g(x)}]}$$

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

$$E_p[f(x)] = \int f(x)p(x)dx = \int \frac{f(x)p(x)}{g(x)} g(x)dx = \frac{E_g\left[\frac{f(x)p(x)}{g(x)}\right]}{E_g[\frac{p(x)}{g(x)}]}$$

$$= \frac{E_g\left[\frac{f(x)p(x)}{g(x)}\right]}{E_g[\frac{p(x)}{g(x)}]} = \frac{\frac{1}{N}\sum_{i=1}^{N} f(x)W(x_i)}{\frac{1}{N}\sum_{j=1}^{N} W(x_j)} \approx \sum_{i=1} \widehat{W}(\mathrm{x_i})f(x_i)$$

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

$$E_p[f(x)] = \int f(x)p(x)dx = \int \frac{f(x)p(x)}{g(x)} g(x)dx = \frac{E_g\left[\frac{f(x)p(x)}{g(x)}\right]}{E_g[\frac{p(x)}{g(x)}]}$$

$$= \frac{E_g\left[\frac{f(x)p(x)}{g(x)}\right]}{E_g[\frac{p(x)}{g(x)}]} = \frac{\frac{1}{N}\sum_{i=1}^{N} f(x)W(x_i)}{\frac{1}{N}\sum_{j=1}^{N} W(x_j)} \approx \sum_{i=1} \widehat{W}(\mathrm{x_i})f(x_i)$$

$$W(x_i) = \frac{p(x_i)}{q(x_i)} \qquad \widehat{W}(x_i) = \frac{W(x_i)}{\sum_{j=0}^{N} W(x_j)}$$

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

In POMCPOW:

Target Distribution $\Pr(s_{t+1} \mid b_t, a_t, o_{t+1})$

Proposal Distribution $\Pr(s_{t+1} \mid b_t, a_t)$

$$W(s_{t+1}) = \frac{p(s_{t+1})}{q(s_{t+1})} = \frac{\Pr(o_t \mid s', a_{t-1}, b_{t-1})}{\Pr(o_t \mid a_{t-1}, b_{t-1})}$$

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

In POMCPOW:

$$b_t(s') = Pr(s'|a_{t-1}, o_t, b_{t-1})$$

Target Distribution

$$= \frac{\Pr(o_t|s', a_{t-1}, b_{t-1}) \Pr(s'|a_{t-1}, b_{t-1})}{\Pr(o_t|a_{t-1}, b_{t-1})}$$

Proposal Distribution $\Pr(s_{t+1}\,|b_t, a_t)$

$$W(s_{t+1}) = \frac{p(s_{t+1})}{q(s_{t+1})} = \frac{\Pr(o_t|s', a_{t-1}, b_{t-1})}{\Pr(o_t|a_{t-1}, b_{t-1})}$$

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

In POMCPOW:

Target Distribution $\Pr(s_{t+1} \mid b_t, a_t, o_{t+1})$

Proposal Distribution $\Pr(s_{t+1} \mid b_t, a_t)$

$$W(s_{t+1}) = \frac{p(s_{t+1})}{q(s_{t+1})} = \frac{\Pr(o_t | s', a_{t-1}, b_{t-1})}{\Pr(o_t | a_{t-1}, b_{t-1})}$$

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

$$\tilde{w}_{\mathcal{P}/\mathcal{Q}}(x) \equiv \frac{w_{\mathcal{P}/\mathcal{Q}}(x)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)} \qquad \text{(SN Importance Weight)}$$

$$d_\alpha(\mathcal{P}||\mathcal{Q}) \equiv \mathbb{E}_{x\sim\mathcal{Q}}[w_{\mathcal{P}/\mathcal{Q}}(x)^\alpha] \qquad \text{(Rényi Divergence)}$$

$$\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \equiv \sum_{i=1}^{N} \tilde{w}_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) \qquad \text{(SN Estimator)}$$

Importance Sampling

Constructs an approximation using samples from a proposal distribution, and corrects for the discrepancy between the target and proposal using (importance) weights.

Theorem 1

Theorem 1 (SN d_∞-Concentration Bound). *Let $\mathcal{P}$ and $\mathcal{Q}$ be two probability measures on the measurable space $(\mathcal{X}, \mathcal{F})$ with $\mathcal{P} \ll \mathcal{Q}$ and $d_\infty(\mathcal{P}||\mathcal{Q}) < +\infty$. Let $x_1, \cdots, x_N$ be i.i.d.r.v. sampled from $\mathcal{Q}$, and $f : \mathcal{X} \to \mathbb{R}$ be a bounded Borel function ($\|f\|_\infty < +\infty$). Then, for any $\lambda > 0$ and N large enough such that $\lambda > \|f\|_\infty d_\infty(\mathcal{P}||\mathcal{Q})/\sqrt{N}$, the following bound holds with probability at least $1 - 3\exp(-N \cdot t^2(\lambda, N))$:*

$$|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \leq \lambda$$

where $t(\lambda, N)$ is defined as:

$$t(\lambda, N) \equiv \frac{\lambda}{\|f\|_\infty d_\infty(\mathcal{P}||\mathcal{Q})} - \frac{1}{\sqrt{N}}$$

Proof 1

Learning And Mining from DatA
http://lamda.nju.edu.cn

Lemma1:

$$
\begin{aligned}
\mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)] \geq \lambda\right) &= \mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x) f(x)\right] \geq \lambda\right) \\
&\leq \exp\left(-\frac{2N^2\lambda^2}{\sum_{i=1}^{N} 2(d_\infty(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty)^2}\right) \\
&\leq exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) \equiv \delta \\
\mathbb{P}\left(\left|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)]\right| \geq \lambda\right) &\leq 2\exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) = 2\delta
\end{aligned}
$$

Proof 1

Lemma1:

$$\mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)] \geq \lambda\right) = \mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x)f(x)\right] \geq \lambda\right) \quad (1)$$

$$\leq \exp\left(-\frac{2N^2\lambda^2}{\sum_{i=1}^{N} 2\left(d_\infty(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty\right)^2}\right)$$

$$\leq exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) \equiv \delta$$

$$\mathbb{P}\left(\left|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)]\right| \geq \lambda\right) \leq 2\exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) = 2\delta$$

(1) $\mathbb{E}_{x\sim P}[f(x)] = \int f(x)p(x)dx = \int \frac{p(x)f(x)}{q(x)} q(x) = \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x)f(x)\right]$

Proof 1

Lemma1:

$$\mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)] \geq \lambda\right) = \mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x)f(x)\right] \geq \lambda\right)$$

$$\leq \exp\left(-\frac{2N^2\lambda^2}{\sum_{i=1}^{N} 2(d_\infty(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty)^2}\right)$$

$$\leq exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) \equiv \delta$$

$$\mathbb{P}\left(\left|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)]\right| \geq \lambda\right) \leq 2\exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) = 2\delta$$

Proof 1

Lemma1:

$$\begin{aligned}
\mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)] \geq \lambda\right) &= \mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x) f(x)\right] \geq \lambda\right) \\
&\leq \exp\left(-\frac{2N^2\lambda^2}{\sum_{i=1}^{N} 2(d_\infty(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty)^2}\right) \quad ② \\
&\leq exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) \equiv \delta \\
\mathbb{P}\left(\left|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)]\right| \geq \lambda\right) &\leq 2\exp\left(-\frac{N\lambda^2}{d_\infty^2(\mathcal{P} \parallel \mathcal{Q}) \parallel f \parallel_\infty^2}\right) = 2\delta
\end{aligned}$$

http://lamda.nju.edu.cn

Proof 1

2 $$\hat{\mu}_{\mathcal{P}/\mathcal{Q}} = \frac{1}{N}\sum_{i=1}^{N} W_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) = \frac{1}{N}\sum_{i=1}^{N} g(x_i)$$ $$g(x) = W_{\mathcal{P}/\mathcal{Q}}(x) f(x)$$

Proof 1

② $\hat{\mu}_{\mathcal{P}/\mathcal{Q}} = \frac{1}{N}\sum_{i=1}^{N} W_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) = \frac{1}{N}\sum_{i=1}^{N} g(x_i)$ $\quad g(x) = W_{\mathcal{P}/\mathcal{Q}}(x) f(x)$

$\hat{\mu}_{\mathcal{P}/\mathcal{Q}}$ is unbiased , which means $E[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}] = E[g(x)]$, then:

Proof 1

② $$\hat{\mu}_{\mathcal{P}/\mathcal{Q}} = \frac{1}{N}\sum_{i=1}^{N} W_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) = \frac{1}{N}\sum_{i=1}^{N} g(x_i) \qquad g(x) = W_{\mathcal{P}/\mathcal{Q}}(x) f(x)$$

$\hat{\mu}_{\mathcal{P}/\mathcal{Q}}$ is unbiased , which means $E[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}] = E[g(x)]$, then:

$$\mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x) f(x)\right] \geq \lambda\right) = \mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}\right] \geq \lambda\right)$$

Proof 1

② $$\hat{\mu}_{\mathcal{P}/\mathcal{Q}} = \frac{1}{N}\sum_{i=1}^{N} W_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) = \frac{1}{N}\sum_{i=1}^{N} g(x_i) \qquad g(x) = W_{\mathcal{P}/\mathcal{Q}}(x) f(x)$$

$\hat{\mu}_{\mathcal{P}/\mathcal{Q}}$ is unbiased , which means $E[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}] = E[g(x)]$, then:

$$\mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x) f(x)\right] \geq \lambda\right) = \mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}\right] \geq \lambda\right)$$

Hoeffding's inequality:

Let $\{X_1, X_2, \dots, X_N\}$ be independent random variables bounded by the interval [a, b], and $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, then:

Proof 1

② $$\hat{\mu}_{\mathcal{P}/\mathcal{Q}} = \frac{1}{N}\sum_{i=1}^{N} W_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) = \frac{1}{N}\sum_{i=1}^{N} g(x_i) \qquad g(x) = W_{\mathcal{P}/\mathcal{Q}}(x) f(x)$$

$\hat{\mu}_{\mathcal{P}/\mathcal{Q}}$ is unbiased , which means $E[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}] = E[g(x)]$, then:

$$\mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[W_{\mathcal{P}/\mathcal{Q}}(x) f(x)\right] \geq \lambda\right) = \mathbb{P}\left(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\left[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}\right] \geq \lambda\right)$$

Hoeffding's inequality:

Let $\{X_1, X_2, \dots, X_N\}$ be independent random variables bounded by the interval [a, b], and $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, then:

$$\forall\, \lambda > 0, \qquad \mathbb{P}\left(\overline{X} - \mathbb{E}_{x\sim Q}\left[\overline{X}\right] \geq \lambda\right) \leq \exp\left(-\frac{2N\lambda^2}{\sum_{i=1}^{N}(b_i - a_i)^2}\right)$$

Proof 1

LAMDA Learning And Mining from DatA http://lamda.nju.edu.cn

2 $$\hat{\mu}_{\mathcal{P}/\mathcal{Q}} = \frac{1}{N}\sum_{i=1}^{N} W_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) = \frac{1}{N}\sum_{i=1}^{N} g(x_i) \qquad g(x) = W_{\mathcal{P}/\mathcal{Q}}(x) f(x)$$

$\hat{\mu}_{\mathcal{P}/\mathcal{Q}}$ is unbiased , which means $E[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}] = E[g(x)]$, then:

$$\mathbb{P}\big(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\big[W_{\mathcal{P}/\mathcal{Q}}(x) f(x)\big] \geq \lambda\big) = \mathbb{P}\big(\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim Q}\big[\hat{\mu}_{\mathcal{P}/\mathcal{Q}}\big] \geq \lambda\big)$$

$$0 \leq g(x_i) \leq \big\|g(x)\big\|_{\infty} = d_{\infty}(\mathcal{P}\|\mathcal{Q})\big\|f(x)\big\|_{\infty}$$

$$\forall\, \lambda > 0, \qquad \mathbb{P}\big(\overline{X} - \mathbb{E}_{x\sim Q}\big[\overline{X}\big] \geq \lambda\big) \leq \exp\Big(-\frac{2N\lambda^2}{\sum_{i=1}^{N}(b_i - a_i)^2}\Big)$$

Proof 1

Learning And Mining from DatA
http://lamda.nju.edu.cn

Further:

$$
\begin{aligned}
&\mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{P}}[f(x)] \geq \lambda) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \geq \tilde{\lambda}) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) \\
&\quad \leq \tilde{\delta} + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda)
\end{aligned}
$$

Proof 1

Further:

$$\begin{aligned}
&\mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{P}}[f(x)] \geq \lambda) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) \qquad ① \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \geq \tilde{\lambda}) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) \\
&\quad \leq \tilde{\delta} + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda)
\end{aligned}$$

① $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$

Proof 1

Further:

$$
\begin{aligned}
&\mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{P}}[f(x)] \geq \lambda) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && 1 \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \geq \tilde{\lambda}) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) && 2 \\
&\quad \leq \tilde{\delta} + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda)
\end{aligned}
$$

2 $$\mathbb{P}\left(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)] \geq \lambda\right)$$

$$= \mathbb{P}\left(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)] - \left|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}\left[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}\right]\right| \geq \tilde{\lambda}\right)$$

$$\leq \mathbb{P}\left(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim P}[f(x)] - (\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}\left[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}\right]) \geq \tilde{\lambda}\right)$$

Proof 1

Learning And Mining from DatA
http://lamda.nju.edu.cn

Further:

$$
\begin{aligned}
&\mathbb{P}(|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{P}}[f(x)] \geq \lambda) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && ① \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \geq \tilde{\lambda}) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) && ② \\
&\quad \leq \tilde{\delta} + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && ③
\end{aligned}
$$

③ Using Lemma1

$$\tilde{\lambda} = \lambda - \left|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]\right|$$

$$\tilde{\delta} = \exp\left(-\frac{N\tilde{\lambda}^2}{d_\infty^2(\mathcal{P}\parallel\mathcal{Q})\parallel f\parallel_\infty^2}\right)$$

Proof 1

Further:

$$
\begin{aligned}
&\mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{P}}[f(x)] \geq \lambda) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && \text{1} \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \geq \tilde{\lambda}) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) && \text{2} \\
&\quad \leq \tilde{\delta} + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && \text{3}
\end{aligned}
$$

Proof 1

Learning And Mining from DatA
http://lamda.nju.edu.cn

Further:

$$\begin{aligned}
&\mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{P}}[f(x)] \geq \lambda) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && 1 \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \geq \tilde{\lambda}) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) && 2 \\
&\quad \leq \tilde{\delta} + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && 3
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) &\leq \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) \\
&\leq \mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \tilde{\lambda}) \leq 2\tilde{\delta}
\end{aligned}$$

Proof 1

Learning And Mining from DatA
http://lamda.nju.edu.cn

Further:

$$
\begin{aligned}
&\mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{P}}[f(x)] \geq \lambda) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && (1) \\
&\quad \leq \mathbb{P}(\tilde{\mu}_{\mathcal{P}/\mathcal{Q}} - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \geq \tilde{\lambda}) + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) && (2) \\
&\quad \leq \tilde{\delta} + \mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) && (3)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \lambda) &\leq \mathbb{P}(\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}} \geq \tilde{\lambda}) \\
&\leq \mathbb{P}(|\,\mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \tilde{\lambda}) \leq 2\tilde{\delta}
\end{aligned}
$$

$$
\mathbb{P}\left(|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \geq \lambda\right) \leq 3\tilde{\delta}
$$

Proof 1

$$
\begin{aligned}
|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]| &= \left|\mathbb{E}_{x\sim\mathcal{Q}}[\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]\right| \leq \mathbb{E}_{x\sim\mathcal{Q}}[|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}|] \\
&\leq \mathbb{E}_{x\sim\mathcal{Q}}\left|\frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)} - \frac{1}{N}\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)\right| \\
&= \mathbb{E}_{x\sim\mathcal{Q}}\left[\left|\frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right|\left|1 - \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right|\right] \\
&\leq \mathbb{E}_{x\sim\mathcal{Q}}\left[\left(\frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right)^2\right]^{1/2} \mathbb{E}_{x\sim\mathcal{Q}}\left[\left(1 - \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2\right]^{1/2} \\
&\leq \|f\|_\infty \sqrt{\frac{d_2(\mathcal{P}||\mathcal{Q}) - 1}{N}} \leq \|f\|_\infty \frac{d_\infty(\mathcal{P}||\mathcal{Q})}{\sqrt{N}}
\end{aligned}
$$

Proof 1

$$
\begin{aligned}
|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]| &= \left|\mathbb{E}_{x\sim\mathcal{Q}}[\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]\right| \leq \mathbb{E}_{x\sim\mathcal{Q}}[|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}|] \\
&\leq \mathbb{E}_{x\sim\mathcal{Q}}\left|\frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)} - \frac{1}{N}\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)\right| \\
&= \mathbb{E}_{x\sim\mathcal{Q}}\left[\left|\frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right|\left|1 - \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right|\right] \qquad 1 \\
&\leq \mathbb{E}_{x\sim\mathcal{Q}}\left[\left(\frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right)^2\right]^{1/2} \mathbb{E}_{x\sim\mathcal{Q}}\left[\left(1 - \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2\right]^{1/2} \\
&\leq \|f\|_\infty \sqrt{\frac{d_2(\mathcal{P}||\mathcal{Q}) - 1}{N}} \leq \|f\|_\infty \frac{d_\infty(\mathcal{P}||\mathcal{Q})}{\sqrt{N}}
\end{aligned}
$$

1 Cauchy-Schwarz inequality

$$
\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)
$$

Proof 1

$$\left| \mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \right| = \left| \mathbb{E}_{x\sim\mathcal{Q}}[\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}] \right| \leq \mathbb{E}_{x\sim\mathcal{Q}}[|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}|]$$

$$\leq \mathbb{E}_{x\sim\mathcal{Q}} \left| \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)} - \frac{1}{N}\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i) \right|$$

$$= \mathbb{E}_{x\sim\mathcal{Q}} \left[\left| \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)} \right| \left| 1 - \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N} \right| \right] \qquad 1$$

$$\leq \mathbb{E}_{x\sim\mathcal{Q}} \left[\left(\frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)} \right)^2 \right]^{1/2} \mathbb{E}_{x\sim\mathcal{Q}} \left[\left(1 - \frac{\sum_{i=1}^{N} w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N} \right)^2 \right]^{1/2} \qquad 2$$

$$\leq \|f\|_\infty \sqrt{\frac{d_2(\mathcal{P}||\mathcal{Q}) - 1}{N}} \leq \|f\|_\infty \frac{d_\infty(\mathcal{P}||\mathcal{Q})}{\sqrt{N}}$$

Proof 1

$$
\begin{aligned}
|\,\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]| &= \left|\mathbb{E}_{x\sim\mathcal{Q}}[\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]\right| \leq \mathbb{E}_{x\sim\mathcal{Q}}[|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}|] \\
&\leq \mathbb{E}_{x\sim\mathcal{Q}}\left|\frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)} - \frac{1}{N}\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)\right| \\
&= \mathbb{E}_{x\sim\mathcal{Q}}\left[\left|\frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right|\left|1 - \frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right|\right] \qquad (1) \\
&\leq \mathbb{E}_{x\sim\mathcal{Q}}\left[\left(\frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i) f(x_i)}{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right)^2\right]^{1/2} \mathbb{E}_{x\sim\mathcal{Q}}\left[\left(1 - \frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2\right]^{1/2} \qquad (2) \\
&\leq \|f\|_\infty \sqrt{\frac{d_2(\mathcal{P}||\mathcal{Q}) - 1}{N}} \leq \|f\|_\infty \frac{d_\infty(\mathcal{P}||\mathcal{Q})}{\sqrt{N}} \qquad (3)
\end{aligned}
$$

Proof 1

$$|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]| = \left|\mathbb{E}_{x\sim\mathcal{Q}}[\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}]\right| \leq \mathbb{E}_{x\sim\mathcal{Q}}[|\hat{\mu}_{\mathcal{P}/\mathcal{Q}} - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}|]$$

$$\leq \mathbb{E}_{x\sim\mathcal{Q}}\left|\frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)}{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)} - \frac{1}{N}\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)\right|$$

$$= \mathbb{E}_{x\sim\mathcal{Q}}\left[\left|\frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)}{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right|\left|1 - \frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right|\right] \quad ①$$

$$\leq \mathbb{E}_{x\sim\mathcal{Q}}\left[\left(\frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)f(x_i)}{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}\right)^2\right]^{1/2}\mathbb{E}_{x\sim\mathcal{Q}}\left[\left(1 - \frac{\sum_{i=1}^N w_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2\right]^{1/2} \quad ②$$

$$\leq \|f\|_\infty\sqrt{\frac{d_2(\mathcal{P}||\mathcal{Q})-1}{N}} \leq \|f\|_\infty\frac{d_\infty(\mathcal{P}||\mathcal{Q})}{\sqrt{N}} \quad ③$$

③ $$d_2(\mathcal{P}||\mathcal{Q}) - 1 \leq d_2(\mathcal{P}||\mathcal{Q}) = \mathbb{E}_{x\sim\mathcal{Q}}[W_{\mathcal{P}/\mathcal{Q}}(x)^2]$$

$$\leq W_{\mathcal{P}/\mathcal{Q}}(x)^2_{max} = d_\infty(\mathcal{P}||\mathcal{Q})^2$$

Proof 1

② $$\mathbb{E}_{x\sim\mathcal{Q}}\left[\left(1-\frac{\sum_{i=1}^{N}W_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2\right]=\mathbb{E}_{x\sim\mathcal{Q}}\left[\sum_{i=1}^{N}\left(\frac{1-W_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2\right]$$

$$\leq\sum_{i=1}^{N}\mathbb{E}_{x\sim\mathcal{Q}}\left[\left(\frac{1-W_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2\right]\leq\sum_{i=1}^{N}\mathbb{E}_{x\sim\mathcal{Q}}\left[\left(\frac{W_{\mathcal{P}/\mathcal{Q}}(x_i)}{N}\right)^2-\left(\frac{1}{N}\right)^2\right]$$

$$=\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{x\sim\mathcal{Q}}\left[\left(W_{\mathcal{P}/\mathcal{Q}}(x_i)\right)^2\right]-\frac{1}{N}\leq\frac{d_2(\mathcal{P}||\mathcal{Q})-1}{N}$$

Proof 1

Last:

$$\tilde{\delta} \leq \exp\left(-\frac{N(\lambda - \|f\|_\infty d_\infty(\mathcal{P}\|\mathcal{Q})/\sqrt{N})^2}{d_\infty^2(\mathcal{P}\|\mathcal{Q})\|f\|_\infty^2}\right)$$

$$= \exp\left(-N\left(\frac{\lambda - \|f\|_\infty d_\infty(\mathcal{P}\|\mathcal{Q})/\sqrt{N}}{\|f\|_\infty d_\infty(\mathcal{P}\|\mathcal{Q})}\right)^2\right)$$

$$\equiv \exp\left(-N \cdot t^2(\lambda, N)\right)$$

Lemma 1

Lemma 1 (SN Estimator Leaf Node Convergence). *$\hat{Q}^*_{D-1}(\bar{b}_{D-1}, a)$ is an SN estimator of $Q^*_{D-1}(b_{D-1}, a)$, and the following leaf-node concentration bound holds with probability at least* $1 - 3\exp(-C \cdot t^2_{\max}(\lambda, C))$,

$$|Q^*_{D-1}(b_{D-1}, a) - \hat{Q}^*_{D-1}(\bar{b}_{D-1}, a)| \leq \lambda$$

Proof 2

$\hat{Q}^*_{D-1}\left(\overline{b}_{D-1}, a\right)$ is an SN estimator of $Q^*_{D-1}\left(\overline{b}_{D-1}, a\right)$:

$$\hat{Q}^*_{D-1}\left(\overline{b}_{D-1}, a\right) = \sum_{i=1}^{C} \widetilde{w}_{\mathcal{P}^{D-1}/\mathcal{Q}^{D-1}}(\{s_n\}_i) R\left(s_{D-1,i}, a\right)$$

$$Q^*_{D-1}\left(b_{D-1}, a\right) = \int_S R(S_{D-1}, a) dS$$

Proof 2

$\hat{Q}^*_{D-1}(\overline{b}_{D-1}, a)$ is an SN estimator of $Q^*_{D-1}(\overline{b}_{D-1}, a)$:

Using Theorem 1:

first bound R by $3V_{max}$:

$$\mathrm{V}_{max} \equiv \frac{R_{max}}{1-\gamma} \geq R_{max}$$

Proof 2

Theorem 1 (SN d_∞-Concentration Bound). *Let $\mathcal{P}$ and $\mathcal{Q}$ be two probability measures on the measurable space $(\mathcal{X}, \mathcal{F})$ with $\mathcal{P} \ll \mathcal{Q}$ and $d_\infty(\mathcal{P}||\mathcal{Q}) < +\infty$. Let $x_1, \cdots, x_N$ be i.i.d.r.v. sampled from $\mathcal{Q}$, and $f : \mathcal{X} \to \mathbb{R}$ be a bounded Borel function ($\|f\|_\infty < +\infty$). Then, for any $\lambda > 0$ and N large enough such that $\lambda > \|f\|_\infty d_\infty(\mathcal{P}||\mathcal{Q})/\sqrt{N}$, the following bound holds with probability at least $1 - 3\exp(-N \cdot t^2(\lambda, N))$:*

$$|\mathbb{E}_{x\sim\mathcal{P}}[f(x)] - \tilde{\mu}_{\mathcal{P}/\mathcal{Q}}| \leq \lambda \tag{6}$$

where $t(\lambda, N)$ is defined as:

$$t(\lambda, N) \equiv \frac{\lambda}{\|f\|_\infty d_\infty(\mathcal{P}||\mathcal{Q})} - \frac{1}{\sqrt{N}} \tag{7}$$

Proof 2

$\hat{Q}^*_{D-1}(\overline{b}_{D-1}, a)$ is an SN estimator of $Q^*_{D-1}(\overline{b}_{D-1}, a)$:

Using Theorem 1:

first bound R by $3V_{max}$:

$$\mathrm{V}_{max} \equiv \frac{R_{max}}{1-\gamma} \geq R_{max}$$

Proof 2

$\hat{Q}^*_{D-1}(\overline{b}_{D-1}, a)$ is an SN estimator of $Q^*_{D-1}(\overline{b}_{D-1}, a)$:

Using Theorem 1:

first bound R by $3V_{max}$: $\mathrm{V}_{max} \equiv \frac{R_{max}}{1-\gamma} \geq R_{max}$

$$t_{D-1}(\lambda, C) = \frac{\lambda}{3V_{\max} d_\infty(\mathcal{P}^{D-1} \parallel \mathcal{Q}^{D-1})} - \frac{1}{\sqrt{C}} \geq \frac{\lambda}{3V_{\max} d_\infty^{\max}} - \frac{1}{\sqrt{C}} \equiv t_{\max}(\lambda, C)$$

$\forall\, i < D,$ $d_\infty(\mathcal{P}^i \parallel \mathcal{Q}^i) \leq d_\infty^{\max}$

$$t_{\mathrm{i}}(\lambda, C) \geq t_{\max}(\lambda, C)$$

Lemma 2

Lemma 2 (SN Estimator Step-by-Step Convergence). *$\hat{Q}^*_d(\bar{b}_d, a)$ is an SN estimator of $Q^*_d(b_d, a)$, and for all $d = 0, \cdots, D-1$ and a, the following holds with probability at least* $1 - 3|A|(3|A|C)^D \exp(-C \cdot t^2_{\max})$:

$$|Q^*_d(b_d, a) - \hat{Q}^*_d(\bar{b}_d, a)| \leq \alpha_d$$

$$\alpha_d \equiv \lambda + \gamma \alpha_{d+1}; \ \alpha_{D-1} = \lambda$$

Proof 3

$$|Q_d^*(b_d, a) - \hat{Q}_d^*(\bar{b}_d, a)| \leq \underbrace{\left|\mathbb{E}[R(s_d, a)|b_d] - \frac{\sum_{i=1}^{C} w_{d,i} r_{d,i}}{\sum_{i=1}^{C} w_{d,i}}\right|}_{\text{(A)}} + \gamma \underbrace{\left|\mathbb{E}[V_{d+1}^*(bao)|b_d] - \frac{\sum_{i=1}^{C} w_{d,i} \hat{V}_{d+1}^*(\overline{b_d a o_i})}{\sum_{i=1}^{C} w_{d,i}}\right|}_{\text{(B)}}$$

Proof 3

$$|Q_d^*(b_d, a) - \hat{Q}_d^*(\bar{b}_d, a)| \leq \underbrace{\left| \mathbb{E}[R(s_d, a)|b_d] - \frac{\sum_{i=1}^{C} w_{d,i} r_{d,i}}{\sum_{i=1}^{C} w_{d,i}} \right|}_{\text{(A)}} + \gamma \underbrace{\left| \mathbb{E}[V_{d+1}^*(bao)|b_d] - \frac{\sum_{i=1}^{C} w_{d,i} \hat{V}_{d+1}^*(\overline{b_d a o_i})}{\sum_{i=1}^{C} w_{d,i}} \right|}_{\text{(B)}}$$

$$|A_1 + B_1 - A_2 - B_2| \leq |A_1 - A_2| + |B_1 - B_2|$$

$$\hat{Q}_d^*(\bar{b}_d, a) = \frac{\sum_{i=1}^{C} w_{d,i} \left(r_{d,i} + \gamma \overline{V^*}_{d+1}(\overline{b_d a o_i}) \right)}{\sum_{i=1}^{C} w_{d,i}}$$

$$Q_d^*(b_d, a) = E[R(s_d, a) + \gamma V_{d+1}^*(bao)|b_d]$$

Proof 3

$$|Q_d^*(b_d,a) - \hat{Q}_d^*(\bar{b}_d,a)| \le \underbrace{\left| \mathbb{E}[R(s_d,a)|b_d] - \frac{\sum_{i=1}^C w_{d,i} r_{d,i}}{\sum_{i=1}^C w_{d,i}} \right|}_{\text{(A)}} + \gamma \underbrace{\left| \mathbb{E}[V_{d+1}^*(bao)|b_d] - \frac{\sum_{i=1}^C w_{d,i} \hat{V}_{d+1}^*(\overline{b_d a o_i})}{\sum_{i=1}^C w_{d,i}} \right|}_{\text{(B)}}$$

For part A: bound R by R_{max} and let $\lambda = \frac{R_{max}}{3V_{max}}\lambda$, using Thero1:

$$A \le \frac{R_{max}}{3V_{max}}\lambda \qquad \Pr \ge 1 - \exp(-t_{\max}^2)$$

Proof 3

For part B:

$$(B) \le \underbrace{\left| \mathbb{E}[V^*_{d+1}(bao)|b_d] - \frac{\sum_{i=1}^{C} w_{d,i} \mathbf{V}^*_{d+1}(s_{d,i}, b_d, a)}{\sum_{i=1}^{C} w_{d,i}} \right|}_{\text{Importance sampling error}}$$

$$+ \underbrace{\left| \frac{\sum_{i=1}^{C} w_{d,i} \mathbf{V}^*_{d+1}(s_{d,i}, b_d, a)}{\sum_{i=1}^{C} w_{d,i}} - \frac{\sum_{i=1}^{C} w_{d,i} V^*_{d+1}(b_d a o_i)}{\sum_{i=1}^{C} w_{d,i}} \right|}_{\text{MC next-step integral approximation error}}$$

$$+ \underbrace{\left| \frac{\sum_{i=1}^{C} w_{d,i} V^*_{d+1}(b_d a o_i)}{\sum_{i=1}^{C} w_{d,i}} - \frac{\sum_{i=1}^{C} w_{d,i} \hat{V}^*_{d+1}(\overline{b_d a o_i})}{\sum_{i=1}^{C} w_{d,i}} \right|}_{\text{Function estimation error}}$$

$$\le \frac{1}{3}\lambda + \frac{2}{3\gamma}\lambda + \alpha_{d+1}$$

Proof 3

For Important Sampling error ,

$$\mathbf{V}^*_{d+1}(s_{d,i}, b_d, a) \equiv \int_S \int_O V^*_{d+1}(b_d a o)\mathcal{Z}(o|a, s_{d+1})\mathcal{T}(s_{d+1}|s_{d,i}, a) ds_{d+1} do$$

$$\begin{aligned}\mathbb{E}[V^*_{d+1}(bao)|b_d] &= \int_S \int_S \int_O V^*_{d+1}(b_d a o)(\mathcal{Z}_{d+1})(\mathcal{T}_{d,d+1}) b_d \cdot ds_{d:d+1} do \\ &= \int_S \mathbf{V}^*_{d+1}(s_d, b_d, a) b_d \cdot ds_d \\ &= \frac{\int_{S^{d+1}} \mathbf{V}^*_{d+1}(s_d, b_d, a)(\mathcal{Z}_{1:d})(\mathcal{T}_{1:d}) b_0 ds_{0:d}}{\int_{S^{d+1}} (\mathcal{Z}_{1:d})(\mathcal{T}_{1:d}) b_0 ds_{0:d}}\end{aligned}$$

Using SN inequality: $\left\|V^*_{d+1}(s_d, b_d, a)\right\|_\infty = V_{max}$

$$\mathrm{IS} \le \frac{1}{3}\lambda \qquad \Pr \ge 1 - \exp(-\mathrm{t}^2_{\max})$$

Proof 3

For MC error:

the quantity $V^*_{d+1}(b_d a o_i)$ for a given $(s_{d,i}, b_d, a)()$ is an unbiased 1-sample MC estimate of $\mathbf{V}^*_{d+1}(s_{d,i}, b_d, a)$

$$\Delta_{d+1}(s_{d,i}, b_d, a) \equiv \mathbf{V}^*_{d+1}(s_{d,i}, b_d, a) - V^*_{d+1}(b_d a o_i) \qquad E(\Delta_{d+1}) = 0$$

$$\left| \frac{\sum_{i=1}^{C} w_{d,i} \mathbf{V}^*_{d+1}(s_{d,i}, b_d, a)}{\sum_{i=1}^{C} w_{d,i}} - \frac{\sum_{i=1}^{C} w_{d,i} V^*_{d+1}(b_d a o_i)}{\sum_{i=1}^{C} w_{d,i}} \right|$$

$$= \left| \frac{\sum_{i=1}^{C} w_{d,i} \Delta_{d+1}(s_{d,i}, b_d, a)}{\sum_{i=1}^{C} w_{d,i}} - 0 \right| \leq \frac{2}{3}\lambda \leq \frac{2}{3\gamma}\lambda \qquad ||\Delta_{d+1}||_\infty \leq 2V_{max}$$

Proof 3

For Function Estimation Error:

the third term is bounded by the inductive hypothesis, since each i-th absolute difference of the Q-function and its estimate at step $d+1$, and furthermore the value function and its estimate at step $d+1$, are all bounded by α_{d+1}.

Proof 3

For part B:

$$|Q_d^*(b_d, a) - \hat{Q}_d^*(\bar{b}_d, a)| \leq \frac{R_{\max}}{3V_{\max}}\lambda + \gamma[\frac{1}{3}\lambda + \frac{2}{3\gamma}\lambda + \alpha_{d+1}]$$

$$\leq \lambda + \gamma\alpha_{d+1} = \alpha_d$$

Thanks!