Given π. and data: $(s_1, a_1, r_1, \cdots, s_H, a_H, r_H)$

we want to evaluate.

"target/eval policy".

$a_{1:H} \sim \pi_b$

"behavior/logging policy".

Importance Sampling/(re)weighting

Inverse Propensity Score (IPS).

$\mathcal{X}$ large finite space.

$p, q \in \Delta(\mathcal{X})$. $f: \mathcal{X} \to [0,1]$. → known.

Want to: estimate $\mathbb{E}_p[f] := \mathbb{E}_{x\sim p}[f(x)]$.

$X_1, X_2, \cdots, X_n \sim p, \Rightarrow \frac{1}{n}\sum_{i=1}^{n} f(X_i)$

"Monte Carlo".

What if: $X_1, X_2, \cdots, X_n \sim q$.

Estimator: $\frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)}{q(X_i)} f(X_i)$

Imp. weight/

density ratio / IPS

"Estimator:" $\boxed{\frac{p(x)}{q(x)} f(x)}$, where $x \sim q$.

Unbiased: $\mathbb{E}_q\left[\frac{p(x)}{q(x)} f(x)\right] = \mathbb{E}_p[f]$.

$$= \sum_x \cancel{q(x)} \frac{p(x)}{\cancel{q(x)}} f(x) = \mathbb{E}_p[f].$$

1. $\|\frac{p}{q}\|_\infty < +\infty$.

$\frac{\|\frac{p}{q}\|_\infty}{\varepsilon^2}$.

2: Hoeffding: $\|\frac{p}{q}\|_\infty \sqrt{\frac{1}{2n} \ln\frac{2}{\delta}}$.

3: $\mathbb{E}_q\left[\frac{p(x)}{q(x)}\right] = \sum_x \cancel{q(x)} \frac{p(x)}{\cancel{q(x)}} = 1$.

4: $\frac{p(x)}{q(x)} f(x) \in [0, \|\frac{p}{q}\|_\infty]$.

$\mathrm{Var}(\quad) = O(\|\frac{p}{q}\|_\infty)$ $X \in [0, a]$.

$\mathrm{Var} X \le O(a^2)$.

$> \frac{p}{q}$.

$\frac{p(x)}{q(x)}$ is large.