

# D-25 密度比推定を用いた逆強化学習

内部英治 銅谷賢治 (OIST)

- 線形可解マルコフ決定過程における関係式

$$\frac{\pi(x' | x)}{p(x' | x)} = \exp(-q(x) - \gamma V(x') + V(x))$$

- $p(x' | x)$ : 学習前の状態遷移.  $\pi(x' | x)$ : 学習後の状態遷移
- $V(x)$ : 価値関数.  $q(x)$ : コスト.  $\gamma$ : 割引率
- 密度比推定を使ったコストと価値の推定
- ヒトの倒立振子の安定化課題に適用

