CME241 A4

Mannual Value Iteration

See assignment4.pdf(written)

Frog Corak

Policy Iterations take longer time to converge than value iterations since for policy iterations policy evaluation and policy improvement alternate until convergence. For example, for 10 pads, value iterations take 1098 steps and policy iterations take 1179 steps.
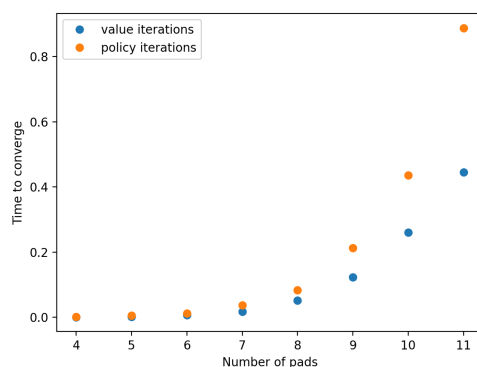

Figure 1: Time to converge

Job Hopping and Wages Utility Maximization

Mathematical Notations

states = {(status, job): status $\in$ {'Employed', 'Unemployed'}, job $\in$ {1,...,n}}
actions = {'Accept', 'Decline'}
state transitions:
$P((\text{status}', i')|(\text{status}, i), \text{Accept}) =$

$$\begin{cases} p'_i\alpha & i' = 1, ..., n, \text{status'} = \text{Unemployed}, \text{status} \in \{\text{Unemployed, Employed}\} \\ 1 - \alpha & i' = i, \text{status'} = \text{Employed}, \text{status} \in \{\text{Unemployed, Employed}\} \\ 0 & \text{otherwise} \end{cases}$$

$$P((\text{status'}, i')|(\text{status}, i), \text{Decline}) =$$
$$\begin{cases} p'_i & \text{status'} = \text{status} = \text{Unemployed}, \\ 0 & \text{otherwise} \end{cases}$$

rewards:
$R((\text{status}, i), \text{Accept} = \log(w_i), \text{status} = \text{Employed}, \text{Unemployed}$
$R((\text{Unemployed}, i), \text{Decline} = \log(w_0)$

Bellman:
$V^*((\text{Unemployed}, i)) = \max\{\log(w_0) + \gamma \sum_{i'=1}^{n} p'_i V^*((\text{Unemployed}, i')), V^*((\text{Employed}, i))\}$
$V^*((\text{Employed}, i)) = \log(w_i) + \gamma(1-\alpha)V^*((\text{Employed}, i)) + \alpha\gamma \sum_{i'=1}^{n} p'_i V^*((\text{Unemployed}, i'))$