1 Introduction

[[Sequential decision problem approximation by simulating possible futures]]
[[Metareasoning to select simulations]]
[[Monte-Carlo tree search]]
[[UCT, which is based on the asymptotically regret-optimal bandit control rule UCB]]
[[Bandit problems vs. selection problems]]
[[Resulting problems for UCT: biased against selecting negative options, no natural stopping criteria]]
[[explain nature of our results: basic theoretical foundations, initial foray into new hueristics for selection and some empirical results]]

2 On optimal policies for selection

[[Formal definition of selection problems and the metalevel MDP with cost per sample (time value); also mention budgeted learning.]]
[[Theorem: if Optimal stops in x, myopic stops in x (converse is more useful!)]]
[[Theorem: if Myopic stops in all states reachable from x, optimal stops in x]]
[[Theorem: expected number of samples is bounded; actual number bounded wp 1]]
[[Theorems: actual number of samples bounded for flat Normal, Bernoulli]]
[[Counterexample to actual boundedness in general (SPRT case)]]

3 Context effects and non-indexability

[[No index theorem; via context inversion counter-example]]
[[Theorem: context effect occurs only for a single convex interval of context value (how general can we make this?]]

4 Regret bounds and approximate policies

[[Regret models: simple regret, regret with cost per sampling; regret goes to zero as c does]]
[[Expected simple regret bounds for normal case?]]
[[Blinkered sampling]]
[[ESPb: Frazier's continuous time approximation]]
[[Concentration upper bounds on VOI]]
[[Control rules: VCT, ECT, BCT]]

5 Application to Monte Carlo tree search

[[summarize MCTS]]

[[Root rules vs. subtree rules; here just focus on former.]]

[[Efficiency gains from reusing samples between time steps; Shimony's model of this]]

[[Simple regret in bandit problem [VEB]CT vs. UCB]]

[[Improvement in performance in Go (or other trees) [VEB]CT vs. UCT]]