The Gittins Index: A Design Principle for Decision-Making Under Uncertainty

Ziv Scully and Alexander Terenin
Cornell University

Abstract

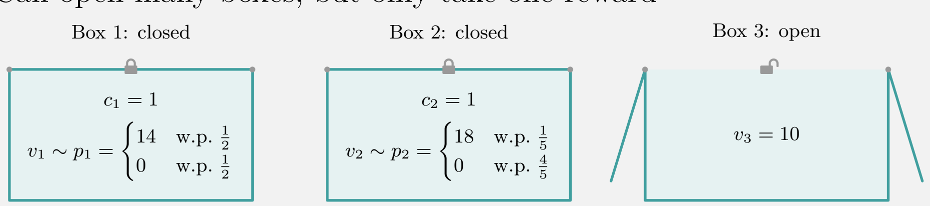
The Gittins index is a tool that optimally solves a variety of decision-making problems involving uncertainty, including multi-armed bandit problems, minimizing mean latency in queues, and search problems like the Pandora's box model. However, despite the above examples and later extensions thereof, the space of problems that the Gittins index can solve perfectly optimally is limited, and its definition is rather subtle compared to those of other multi-armed bandit algorithms. As a result, the Gittins index is often regarded as being primarily a concept of theoretical importance, rather than a practical tool for solving decision-making problems.The aim of this tutorial is to demonstrate that the Gittins index can be fruitfully applied to practical problems. We start by giving an example-driven introduction to the Gittins index, then walk through several examples of problems it solves - some optimally, some suboptimally but still with excellent performance. Two practical highlights in the latter category are applying the Gittins index to Bayesian optimization, and applying the Gittins index to minimizing tail latency in queues.

Pandora's Box

Decision problem with collection of boxes with unknown rewards:

- Rewards: $v(x_i) \sim p_i$ (where p_i is known)
- Cost to open: $c(x_i)$
- Goal: maximize $\mathbb{E} \max_{1 \le t \le T} v(x_t) - \mathbb{E} \sum_{t=1}^{T} c(x_t)$

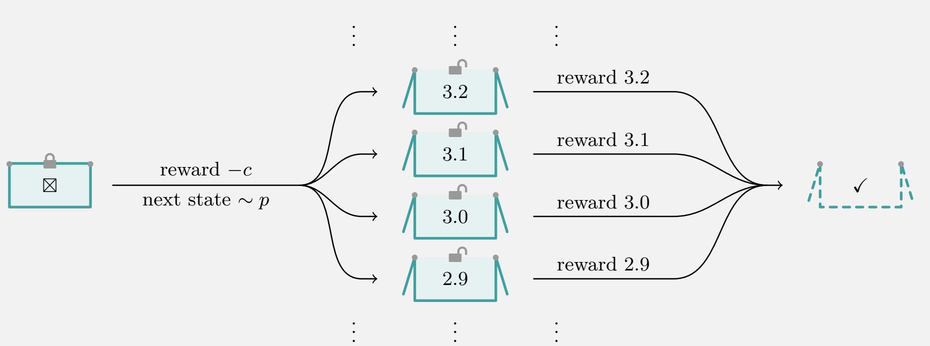
Can open many boxes, but only take one reward



Optimal policy: (notation: $\mathrm{EI}_\psi(x; y) = \mathbb{E}\max(0, \psi(x) - y)$)

Maximize $\alpha^\star(x) = g$ where g solves $\mathrm{EI}_f(x; g) = c(x)$

Pandora's Box as a Markov Chain



Transitive Markov chain over $S = \{\boxtimes\} \cup \mathbb{R} \cup \{\checkmark\}$

Markov Chain Selection

Abstract decision problem:

- Given: collection of (independent) Markov chains, each with states S_i, terminal states ∂S_i transition kernel p_i, reward r_i
- Actions: at each time, choose to transition one Markov chain
- Goal: maximize expected total rewards

Generalizes Pandora's box: also admits an explicit optimal policy!

Local MDPs

Key decision problem difficulty: Markov chains behave randomly

- *Idea*: compare each Markov chain with deterministic number α

Definition. *Let $(S, \partial S, p, r)$ be a Markov chain. For every* ALTERNATIVE OPTION *$\alpha \in \mathbb{R}$ and* INITIAL STATE *$s \in S$, define a Markov decision process, called the (s, α)-*LOCAL *MDP, as follows:*

1. *State space: let $S_{\text{loc}} = S \cup \{\checkmark\}$, with initial state s.*
2. *Terminal state: $\partial S_{\text{loc}} = \{\checkmark\}$.*
3. *Action space: let $A_{\text{loc}} = \{\rhd, \square\}$, called* GO *and* STOP.
4. *Reward function: for $s \in S$, let $r_{\text{loc}}(s, \rhd) = r(s)$, $r_{\text{loc}}(s, \square) = \alpha$, and $r_{\text{loc}}(\checkmark, \rhd) = r_{\text{loc}}(\checkmark, \square) = 0$.*
5. *Transition kernel: if $s \in S$ and $a = \rhd$, then let $s' \sim p(s)$, otherwise if $s = \checkmark$ or $a = \square$ let $s' = \checkmark$.*

Optimal policy: tells us whether to prefer the Markov chain, or α

The Gittins Index

Idea: what if $\square$ and $\rhd$ are co-optimal?

Definition. *Let $(S, \partial S, p, r)$ be a Markov chain. The* GITTINS INDEX, *denoted $G : S \to \mathbb{R} \cup \{\infty\}$, maps each state $s \in S$ to either the unique number $g \in \mathbb{R}$ such that both $\rhd$ and $\square$ are optimal actions for the (g, s)-local MDP at its initial state, or to ∞ if no such number exists.*

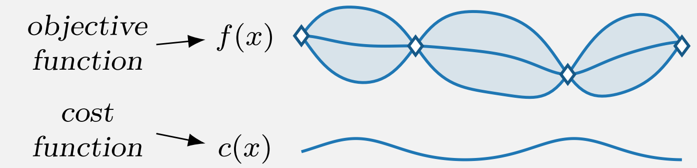
Theorem. *A policy for Markov chain selection is optimal if and only if it selects an action of maximal Gittins index, namely*

$$a \in \operatorname*{arg\,max}_{i \in \{1,\ldots,n\}} G_i(s_i).$$

Maximizing Gittins index: optimal for Markov chain selection

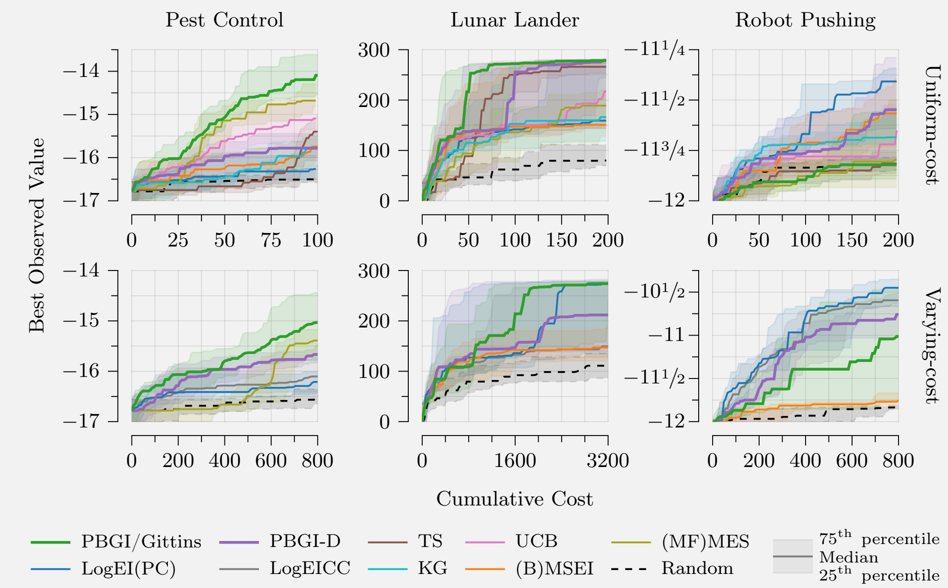
- Often empirically strong for harder decision problems

Cost-aware Bayesian Optimization



Like Pandora's Box, but with correlations:

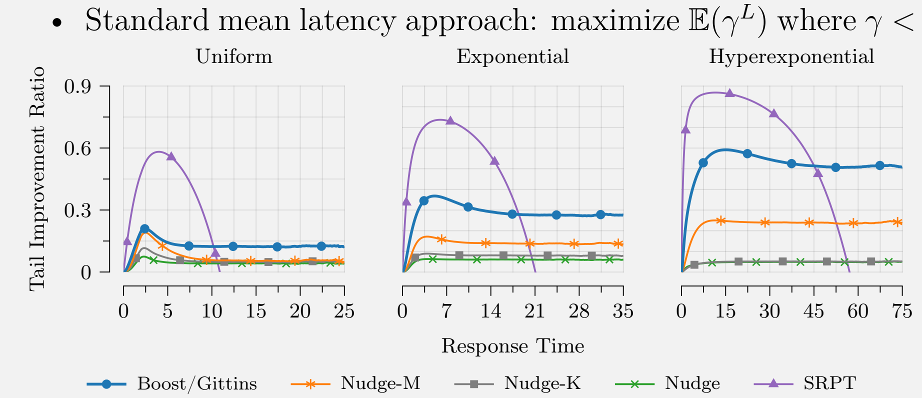
- *Idea*: Gittins Index using reward distribution given by posterior
- Results in empirically strong policy, rather than an optimal one



Tail Scheduling for Queues

Objective: minimize tail latency $\mathbb{P}(L > t)$ as $t \to \infty$

- Standard mean latency approach: maximize $\mathbb{E}(\gamma^L)$ where $\gamma < 1$



Key idea: Gittins index variant with *inflation* rather than discounting

- Tail scheduling approach: minimize $\mathbb{E}(\gamma^L)$ where $\gamma > 1$