Gittins Index Tutorial

Overview

What is the Gittins index?

Solution to the Bayesian/Markovian multi-armed bandit problem

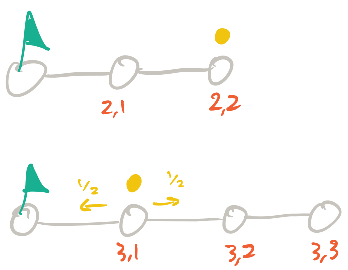
How should one understand the Gittins index? | What other problems does the Gittins index solve?

- $\propto$ super slick proof → combinatorial optimization
- ΔV dynamic programming → "superprocesses"
- queueing theory glasses → arrivals, long-run avg cost

First: golf!

Intro

Markov-process multi-golf



Multi-golf:

- Every turn, hit one ball
- When hit, ball moves to random neighbor
- Goal: minimize $\underbrace{\text{time (\# hits)}}_{\Phi^\pi}$ to get either ball to flag

↳ $\min_{\text{policy } \pi} E[\Phi^\pi]$

Question: Which ball to hit first?

Use expected # hits to score one ball?

Let S_x = time from • at x to flag (if not interrupted)

("swings")

$E[S_{2,1}] = 1 + \frac{1}{2}E[S_{2,2}] = 3$

$E[S_{2,2}] = 1 + E[S_{2,1}] = 4$

$E[S_{3,1}] = 5$

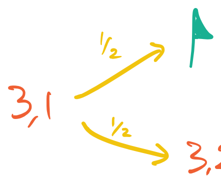
$E[S_{3,2}] = 8$

$E[S_{3,3}] = 9$

$E[S_{2,2}] < E[S_{3,1}]$

... but should hit 3,1 first!

Why 3,1 before 2,2?



3,2 ⇒ switch to 2,2 ball

Extreme case: ∞,1 vs 2,2

$\infty,1 < 2,2 < \infty,2$

"priority ordering"

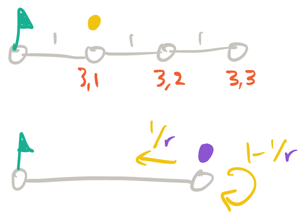
Takeaway:

- $E[S_x]$ is relevant if we would hit • from x to ⚑ without interruption
- But we might interrupt if another • is better

⇒ $E[S_x]$ is not enough!

Generalization: any Markov chain/process, any cost/time per edge

Key idea: Calibrate each state x to a simple benchmark



Pay-to-win golf
- Every turn, hit ball • (just one)
  OR give up by paying r
- Goal: minimize total cost to get a ⚑:

$$\min_{\text{policy } \pi}\Big(E[\text{cost from hitting}] + E[\text{cost from giving up}]\Big)$$

$$= \min_{\text{state sets } Y}\Big(\underbrace{E[\text{time to exit } Y]}_{=S_x(Y)} + r\,\underbrace{P[\text{when out of } Y,\text{ not at ⚑}]}_{=h_x(Y)}\Big)$$

(starting from x) (starting from x)

$E[S_x(Y)] = s_x(Y)$ "had enough"

Def: Pay-to-win golf cost fns:

$$\varphi_x(r,Y) = s_x(Y) + r\,h_x(Y)$$

$$Y^*(r) = \operatorname*{argmax}_Y \varphi_x(r,Y)$$

$$\varphi_x(r) = \varphi_x(r, Y^*(r))$$

$$= \underbrace{s_x(Y^*(r))} + r\,\underbrace{h_x(Y^*(r))}$$

$$= s_x(r) + r\,h_x(r)$$

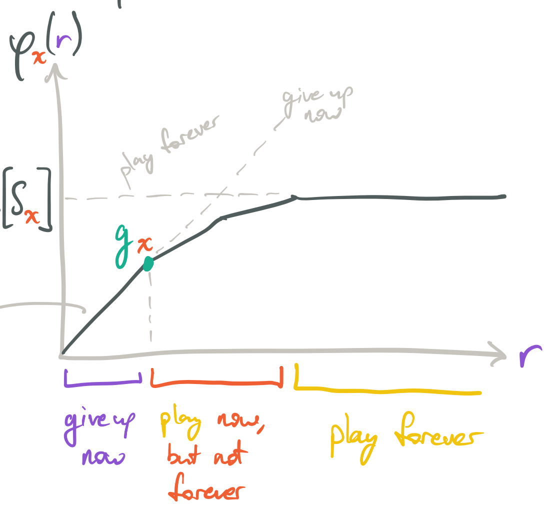
For short → $\varphi = s + rh$

How does pay-to-win golf go?
- Play while in $Y^*(r)$
- Give up if we exit $Y^*(r)$

What does $\varphi_x(r)$ look like? How can we use it to "rate" state x?

$\varphi_x(\infty) = E[S_x]$

$\varphi_x(r) = r$ for small r $(x \notin Y^*(r))$



Question: Where on this curve should I look to get a "rating" for x?

- Play forever line?
- Give up now line?
- In between?

Def: The Gittins rank of state x is:

$$\begin{aligned} g_x &= \inf\{r \geq 0 \mid \varphi_x(r) < r\} \\ &= \inf\{r \geq 0 \mid x \in Y^*(r)\} \\ &= \inf\{r \geq 0 \mid \text{play at least once}\} \end{aligned}$$

(The Gittins index is $\frac{1}{g_x}$)

Exercise: Find $g_{2,1}, g_{2,2}, g_{3,1}, g_{3,2}, g_{3,3}, g_{\infty,1}, g_{\infty,2}, g_{\infty,k}, g_{n,k}$

Hint: In each case, what is $\underbrace{Y^*(g_x)?\ s_x(g_x)?\ h_x(g_x)?}_{\hookrightarrow \text{Include } x \in Y^*(g_x)}$

Lem: $g_x = \inf_Y \frac{s_x(Y)}{1-h_x(Y)} = \frac{s_x(g_x)}{1-h_x(g_x)}$ ← short for $Y^*(g_x)$

Pf: $g_x < r \iff \underbrace{\exists Y,}_{\substack{\hookrightarrow \text{Can pick } Y^*(g_x) \text{ if } g_x < r, \\ \text{optimal as } r \to g_x}} s_x(Y) + r h_x(Y) < r \iff \exists Y, \frac{s_x(Y)}{1-h_x(Y)} < r$

(Going to be less color-coded hereafter...)

Gittins optimality proof I

∝ Super slick proof

Gittins policy always plays ball of <u>minimal</u> g_x

Optimal for one stochastic ball, one deterministic ball (pay-to-win golf)

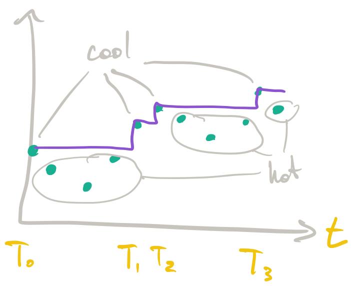
Two stochastic balls?

Key idea: Reduce two-ball case to random sequence of pay-to-win problems

Def: Consider trajectory $X_x(t)$ with $X_x(0) = x$

- The **Gittins envelope** is $G_x(t) = \max_{0 \le u \le t} g_{X(u)}$
- The **Gittins maximum** is $G_x = G(\infty)$
- Ball is **cool** if $G(t) = g_{X(t)}$, else **hot**

(Often omit x subscript)



Consider following sequence of times

- $T_0 = 0$, play through pay-to-win with $r = g_x = G(T_0)$

  We give up iff $G_x > r$, and if so...
- $T_1 = S_x(G(T_0))$ is when we give up

  ⋮
- $T_{i+1} = S_x(G(T_i)) = T_i + S_{X(T_i)}(g_{X(T_i)})$ while $G_x > g_{X(T_i)}$

T_i is i^{th} time ball is cool

Consider $\varphi_x(g_x) = s_x(g_x) + g_x h_x(g_x) = g_x$

by def. of g_x (or lemma)

$$\Rightarrow g_x(1 - h_x(g_x)) = s_x(g_x)$$

$$\Rightarrow E[G_x \mathbb{1}(\text{get to } A)] = E[\text{time spent playing}]$$

Intuition: imagine instead of pay-to-win, we get a reward when done

More generally: $E[G_x \mathbb{1}(\underset{\text{in round } i}{\text{get to } A}) \mid T_i, X(T_i)] = E[\underbrace{G(T_i)}_{=G_i} \mathbb{1}(A_i) \mid T_i, X(T_i)]$

$\parallel$

$E[\text{time played in heat } i \mid \cdots] = E[\underbrace{T_{i+1} - T_i}_{=S_i} \mid T_i, \underbrace{X(T_i)}_{=X_i}]$

Lem: $E[G_i \mathbb{1}(A_i) \mid X_i, T_i] = E[S_i \mid X_i, T_i]$

Def: The **pseudotime** of heat i is $G_i \mathbb{1}(A_i) = G_x \mathbb{1}(A_i)$

So $E[\text{pseudotime of heat}] = E[\text{time of heat}]$

But pseudotime is all at the end of heat, if at all, so

$E[\text{pseudotime of incomplete heat}] = 0 \le E[\text{time of incomplete heat}]$

Lem: For any stopping time τ, $E[\text{pseudotime by } \tau] \le E[\text{time by } \tau] = E[\tau]$

Becomes $=$ if always cool at τ

All of this is still for one ball... but still true for two balls interleaved!

Lem: For any policy π, $E[\overbrace{\text{pseudotime of } \pi}^{=\Psi^\pi}] \le E[\overbrace{\text{time of } \pi}^{=\Phi^\pi}]$

Becomes $=$ if π only interrupts a ball when it's cool

We have:

$E[\Psi^\pi] \le E[\Phi^\pi]$

$E[\Psi^{\text{Gittins}}] = E[\Phi^{\text{Gittins}}]$

Same reasoning for any # of jobs

Suffices:

$E[\Psi^{\text{Gittins}}] \le E[\Psi^\pi]$

Question: Is this true? Why?

- Yes!
- Because $\Psi^\pi \ge \min\{G_{x_1}, G_{x_2}\} = \Psi^{\text{Gittins}}$

Thm: Gittins policy is optimal, w/

$E[\Phi^{\text{Gittins}}_{\vec{x}}] = E[\min_j G_{x_j}]$

Extension: First k balls, set of balls is matroid base, ... [Singla '18, GJSS '19]

Gittins optimality proof 2

ΔV Dynamic programming

Let $\Phi(\vec{x}) = E\left[\Phi^{\text{Gittins}}_{\vec{x}}\right] = E\left[\min_j G_{x_j}\right]$

Φ is cost-to-go fn of multi-golf

Can we show Φ satisfies Bellman eq?

Two steps:

- Write Bellman eq
- Think about G_x to understand Φ

Exercise: What is $P[G_x > r]$?

$G_x > r \iff$ give up in pay-to-win from x w/ $\underbrace{\text{cost } r}$

or time to give up

$\Rightarrow P[G_x > r] = P[\text{give up}] = h_x(r)$

Exercise: What is $E[\min\{G_x, r\}]$?

$\min\{G_x, r\}$ is $\underbrace{\text{pseudotime}}_{G_{\text{give-up}} = r}$ of pay-to-win golf, $\underbrace{\text{playing optimally}}$

So $E[\text{time}] = E[\text{pseudotime}]$
$= \varphi_x(r)$

$\Rightarrow E[\min\{G_x, r\}] = \varphi_x(r) = s_x(r) + r\,h_x(r)$

Exercise: Write $E[\min\{G_{x_1}, G_{x_2}\}]$ in two ways: using h and using φ, G

Hint: $E[R] = \int_0^\infty P[R > r]\,dr$
for non-negative rv R

$= \int_0^\infty h_{x_1}(r)\,h_{x_2}(r)\,dr = E[\varphi_{x_1}(G_{x_2})] = E[\varphi_{x_2}(G_{x_1})]$

Generalizing to any # of balls:

$$\Phi(\vec{x}) = \int_0^\infty \prod_j h_{x_j}(r)\,dr = E\left[\varphi_{x_i}\left(\min_{j\neq i} G_{x_j}\right)\right]$$

Time to show Φ solves Bellman eq

Def:

- $\underbrace{\Delta_i f(x_i)}_{f(\ldots,\, x_i,\, \ldots)} = E\left[f(X_{next}) \mid \text{start at } x_i\right] - f(x_i)$
- $\Gamma f(r) = -\frac{1}{r} f(r) = \frac{1}{r} 0 + \frac{r-1}{r} f(r) - f(r)$

(changing notation: $\varphi_x(r) \rightsquigarrow \varphi(x,r)$)

Bellman eq for pay-to-win golf:

$$\underbrace{\text{“}\min\{\Delta_i, \Gamma\}\varphi(x_i, r)\text{”}}_{\text{change in cost-to-go}} = \underbrace{-1}_{\text{cost paid this round}}$$

Specifically:

$$\Delta_i \varphi(x_i, r) \geq -1,$$

with = iff $x_i \in Y^*(r) \Leftrightarrow g_{x_i} \leq r$

Bellman eq for multi-golf:

$$\min_i \Delta_i \Phi(\vec{x}) = -1$$

Want to show

$$\Delta_i \Phi(\vec{x}) \geq -1,$$

with = iff $g_{x_i} \leq g_{x_j} \quad \forall j \neq i$

$$\Delta_i \Phi(\vec{x}) = E\left[\Delta_i \varphi_{x_i}\left(\min_{j\neq i} G_{x_j}\right)\right] \geq -1$$

with = iff $g_{x_i} \leq G_{x_j}$ w.p. 1 $\Leftrightarrow g_{x_i} \leq g_{x_j}$

because g_{x_j} is min possible value of G_{x_j}

Extension: Sometimes goes through when multiple actions possible per ball ("choice of club")

[Whittle '80, Doval '18]

Gittins optimality proof 3

Queueing theory glasses

Change of objective: minimize <u>sum</u> of completion times $\Leftrightarrow$ minimize avg # incomplete jobs

Question: Can we adapt previous proofs to this setting? Yes!

What if jobs arrive over time? Harder...

Recall: $\frac{\partial}{\partial r}\varphi_x(r) = h_x(r) \Rightarrow \frac{\partial}{\partial r}\frac{\varphi_x(r)}{r} = \frac{-s_x(r)}{r^2}$

$$\Rightarrow \int_0^\infty \underbrace{\frac{s_x(r)}{r^2}dr}_{\text{"}d\frac{1}{r}\text{"}} = \text{"}\frac{\varphi_x(0)}{0}\text{"} - \text{"}\frac{\varphi_x(\infty)}{\infty}\text{"} = 1$$

Lem: $\int_0^\infty E[S_x(r)]\,d\frac{1}{r} = 1$

Let $W_{\vec{x}}(r) = \sum_j S_{x_j}(r)$ be "r-work" in system

Lemma $\Rightarrow \int_0^\infty E[W_{\vec{x}}(r)]\,d\frac{1}{r} = \#\vec{x}$

Let $W(r)$ be r-work at "uniformly random" time

Thm: $\int_0^\infty E[W(r)]\,d\frac{1}{r} = E[\#\text{jobs}]$

WINE: Work Integral Number Equality

Question: How to minimize $W(r)$?

Thm: Gittins minimizes $E[\#\text{jobs}]$

Extension: Still holds w/ Poisson arrivals

[SGH '20, SH '21]