A More General Pandora's Rule

Richard Weber, University of Cambridge

Wojciech Olszewski (Dept of Economics, Northwestern U)

John William Waterhouse: Pandora, 1896

Presented in Management Science Seminar Series,
London School of Economics, 13 November, 2013

A More General Pandora's Rule?

Richard Weber, University of Cambridge

Wojciech Olszewski (Dept of Economics, Northwestern U)

John William Waterhouse: Pandora, 1896

Presented in Management Science Seminar Series,
London School of Economics, 13 November, 2013

Abstract

In a famous model due to Weitzman (1979) an agent (Pandora) is presented with boxes containing prizes. She may open them however as she likes, discover prizes within, and optimally stop. Her aim is to maximize the expected value of the greatest prize she finds, minus the costs of opening boxes. This problem has an attractive solution by means of the so-called Pandora rule, and might be applied to searching for a research topic, house or job.

It does not, however, address the problem of a student who is searching for the subject to choose as her major and who benefits from all the courses she takes, not just from those taken once her major is chosen. So motivated, we set out to discover whether there exist any problems for which a Pandora rule is optimal when the aim is to maximize is a more general function of all the revealed prizes. We elucidate the connection between the Pandora rule and the Gittins index solution of an equivalent multi-armed bandit problem.

Although the Gittins index analysis tells most of the story, there do exist problems which are not equivalent to multi-armed bandits and for which a Pandora rule is optimal. We give a sufficient conditions that can be used to identify this and an example of its application.

Weitzman's Pandora problem

Weitzman's Pandora problem

Econometrica, Vol. 47, No. 3 (May, 1979)

OPTIMAL SEARCH FOR THE BEST ALTERNATIVE

BY MARTIN L. WEITZMAN[1]

This paper completely characterizes the solution to the problem of searching for the best outcome from alternative sources with different properties. The optimal strategy is an elementary reservation price rule, where the reservation prices are easy to calculate and have an intuitive economic interpretation.

Weitzman "OPTIMAL SEARCH FOR THE BEST ALTERNATIVE"

Web Images Maps Shopping More Search tools

Page 25 of about 25,300 results (0.23 seconds)

[PDF] Search Costs as a Means for Improving Market Performance
www.cs.biu.ac.il/~sarned/pub/AMEC2010Sarne.pdf
by D Sarne - Related articles
... volume 1, chapter 13, pages 769–847. Else- vier, 1 edition, 1989. 26. M. L. **Weitzman**. **Optimal search for the best alternative**. Econometrica, 47(3):641–54,.

[PDF] Expert-Mediated Search
www.cs.biu.ac.il/~sarned/pub/AAMAS2011SarneExpert.pdf
by M Chhabra - 2011 - Cited by 8 - Related articles
The. Quarterly Journal of Economics, 102(3):581–593, 1987. [18] M. **Weitzman**. **Optimal search for the best alternative**. Econometrica: Journal of the Econometric ...

[PDF] Exploration and Exploitation During Sequential Search - Bayesia...
klab.smpp.northwestern.edu/.../Dam_Explore_exploit_sequent_search_2...
by G Dam - 2009 - Cited by 6 - Related articles
ing the success of random searches. Nature, 401, 911–914. **Weitzman**, M. L. (1979). **Optimal search for the best alternative**. Econometrica, 47 (3), 641–654.

References - Journal of Travel Research - Sage Publications
jtr.sagepub.com/content/43/4/328.refs
by C Blain - 2005 - Cited by 341 - Related articles
Weitzman, M. L.. **Weitzman**, M. L. (1979). "**Optimal Search for the Best Alternative**." Econometrica , 47: 641-655. CrossRefWeb of Science. Woodside, Arch.

Strategic price complexity in retail financial markets - Ideas - RePEc
ideas.repec.org/a/eee/jfinec/v91y2009i3p278-287.html
by Bl Carlin - 2009 - Cited by 165 - Related articles
M. L. **Weitzman**, 1978. "**Optimal Search for the Best Alternative**," Working papers 214, Massachusetts Institute of Technology (MIT), Department of Economics.

Weitzman's Pandora problem

Martin L. Weitzman is Professor of Economics at Harvard University.

Weitzman's Pandora problem

- Pandora has n boxes.

Weitzman's Pandora problem

- Pandora has n boxes.
- Box i contains a prize, of value x_i, distributed with known c.d.f. F_i.

Weitzman's Pandora problem

- Pandora has n boxes.
- Box i contains a prize, of value x_i, distributed with known c.d.f. F_i.
- At known cost c_i she can open box i and discover x_i^o.

Weitzman's Pandora problem

- Pandora has n boxes.
- Box i contains a prize, of value x_i, distributed with known c.d.f. F_i.
- At known cost c_i she can open box i and discover x_i^o.
- Pandora may open boxes in any order, and stop at will.

Weitzman's Pandora problem

- Pandora has n boxes.
- Box i contains a prize, of value x_i, distributed with known c.d.f. F_i.
- At known cost c_i she can open box i and discover x_i^o.
- Pandora may open boxes in any order, and stop at will..
- She opens a subset of boxes $S \subseteq \{1, \ldots, n\}$ and then stops. She wishes to maximize the expected value of

$$R = \max_{i \in S} x_i - \sum_{i \in S} c_i.$$

Reasons for liking Weitzman's problem

Weitzmans' problem is attractive.

1. It has many applications:
   - hunting for a house
   - selling a house (accepting the best offer)
   - searching for a job,
   - looking for research project to focus upon.

Reasons for liking Weitzman's problem

Weitzmans' problem is attractive.

1. It has many applications:
   - hunting for a house
   - selling a house (accepting the best offer)
   - searching for a job,
   - looking for research project to focus upon.
2. It has an index policy solution, a so-called **Pandora rule**.

Varian's problem: 'economics and search'

Hal Varian (1999) put Weitzman's problem like this:

- You work at airport book store;
- people are in a hurry;
- mental effort to examining books ($c > 0$);
- will only take one book with them;
- you have an idea of how likely it is that person will like the book ($F_i(x)$).

Problem: in what order to show them books?

Varian's problem: 'economics and search'

Hal Varian (1999) put Weitzman's problem like this:

- You work at airport book store;
- people are in a hurry;
- mental effort to examining books ($c > 0$);
- will only take one book with them;
- you have an idea of how likely it is that person will like the book ($F_i(x)$).

Problem: <u>in what order to show them books?</u>

Customer runs in says "I want a travel guide to Borneo."

Which do you show first: Fodors or Lonely Planet?

If only time for one book, show Fodors

If time for two books, show Lonely Planet

Scheduling and stopping

Weitzman's Pandora problem has aspects of both

- **Scheduling**: in what order should the boxes be opened?
- **Stopping**: when should one be content to take the greatest prize found thus far?

Scheduling and stopping

Weitzman's Pandora problem has aspects of both

- **Scheduling**: in what order should the boxes be opened?
- **Stopping**: when should one be content to take the greatest prize found thus far?

Pure scheduling problems often solved by interchange arguments.

Scheduling and stopping

Weitzman's Pandora problem has aspects of both

- **Scheduling**: in what order should the boxes be opened?
- **Stopping**: when should one be content to take the greatest prize found thus far?

Pure scheduling problems often solved by interchange arguments.

Pure stopping problems often solved by one-step-look-ahead rule.

An interchange argument

- Box i is either empty (w.p. q_i) or has €1 (w.p. $p_i = 1 - q_i$)
- Costs c_i to look in box i.
- Wish to minimize expected cost of finding €1.

An interchange argument

- Box i is either empty (w.p. q_i) or has €1 (w.p. $p_i = 1 - q_i$)
- Costs c_i to look in box i.
- Wish to minimize expected cost of finding €1.

Best to search $i, j, \dots$ rather than $j, i, \dots$ if

$$c_i + q_i(c_j + q_j X) < c_j + q_j(c_i + q_i X)$$

i.e. if $c_i/p_i < c_j/p_j$.

So optimal to search in increasing order of **index** c_i/p_i.

Reservation values

Weitzman's problem. A 'reservation price' (or value) for box i is determined by calibration, by asking

"for what value of prize, already available, would we be indifferent between taking that prize, or opening box i and then taking the best prize available?"

$$x_i^* = \inf\Big\{ y : y \geq -c_i + E[\max\{x_i, y\}] \Big\}$$

Reservation values

Weitzman's problem. A 'reservation price' (or value) for box i is determined by calibration, by asking

"for what value of prize, already available, would we be indifferent between taking that prize, or opening box i and then taking the best prize available?"

$$x_i^* = \inf\Big\{y : y \geq -c_i + E[\max\{x_i, y\}]\Big\}$$

$$= \inf\Big\{y : c_i \geq E[\max\{x_i - y, 0\}]\Big\}.$$

Reservation values

Weitzman's problem. A 'reservation price' (or value) for box i is determined by calibration, by asking

"for what value of prize, already available, would we be indifferent between taking that prize, or opening box i and then taking the best prize available?"

$$
\begin{aligned}
x_i^* &= \inf\Big\{ y : y \geq -c_i + E[\max\{x_i, y\}] \Big\} \\
&= \inf\Big\{ y : c_i \geq E[\max\{x_i - y, 0\}] \Big\}.
\end{aligned}
$$

Has the character of a one-step-look-ahead rule.

If best prize found has value $x < x_i^*$ then we should not stop, since

$$
x < -c_i + E[\max\{x_i, x\}].
$$

Pandora rule

Weitzman's Pandora rule.

SELECTION RULE: If a box is to be opened, it should be that closed box with highest reservation price.

STOPPING RULE: Terminate search whenever the maximum sampled reward exceeds the reservation price of every closed box.

Pandora rule

Weitzman's Pandora rule.

SELECTION RULE: If a box is to be opened, it should be that closed box with highest reservation price.

STOPPING RULE: Terminate search whenever the maximum sampled reward exceeds the reservation price of every closed box.

"That such an elementary decision strategy as Pandora's Rule is optimal depends more crucially than might be supposed on the simplifying assumptions of the model. There does not seem to be available a sharp characterization of an optimal solution when certain features of the present model are changed.

Pandora's Rule does not readily generalize." (Weitzman, 1979)

Generalizing Weitzman's problem

In Weitzman's problem the reward is

$$R = \max_{i \in S} x_i - \sum_{i \in S} c_i.$$

Let's try to generalize this, so that if a set of boxes $S \subset \{1, 2, \ldots, n\}$ have been opened:

$$R = u(x_S) - \sum_{i \in S} c_i,$$

where $x_S = (x_i : i \in S)$, and $u(\cdot)$ is some general function of x_S.

Motivating applications

$$R = u(x_S) - \sum_{i \in S} c_i.$$

- Student benefits from the courses she takes while searching for the subject to choose as her major;
- Person obtains a flow utility of dating with different partners in the process of looking for a spouse;
- Organization which experiments with different forms of organization, before adopting a more permanent form, is affected by those temporary forms.

Generalized Pandora rule

Weitzman's Pandora rule.

Open the unopened box with greatest reservation value, until all reservations values are less than the greatest prize that has been found.

Generalized Pandora rule

Weitzman's Pandora rule.

Open the unopened box with greatest reservation value, until all reservations values are less than the greatest prize that has been found.

Generalized Pandora rule.

Open the unopened box with greatest reservation value, until all reservations values are 0 or all boxes have been opened.

Generalized reservation values

Weitzman's problem. The reservation value of box i is

$$x_i^* = \inf\Big\{y : y \geq -c_i + E[\max\{x_i, y\}]\Big\}$$
$$= \inf\Big\{y : c_i \geq E[\max\{x_i - y, 0\}]\Big\}.$$

Generalized reservation values

Weitzman's problem. The reservation value of box i is

$$x_i^* = \inf\Big\{y : y \geq -c_i + E[\max\{x_i, y\}]\Big\}$$
$$= \inf\Big\{y : c_i \geq E[\max\{x_i - y, 0\}]\Big\}.$$

Generalized Weitzman problem. Now the reservation value of box i must depend on what as already been uncovered, x_S.

$$x_i^*(x_S) = \inf\Big\{y : u(x_S, y) \geq -c_i + E[u(x_S, x_i, y)]\Big\}$$

$=$ smallest prize whose addition to prizes already discovered makes it as good to stop as to open box i and then stop.

Generalized utility function

Recall we wish to maximize the expected value of

$$R = u(x_S) - \sum_{i \in S} c_i.$$

For what utility functions u is a Pandora rule optimal?

Generalized utility function

Recall we wish to maximize the expected value of

$$R = u(x_S) - \sum_{i \in S} c_i.$$

For what utility functions u is a Pandora rule optimal?

Generalized utility function

Recall we wish to maximize the expected value of

$$R = u(x_S) - \sum_{i \in S} c_i.$$

For what utility functions u is a Pandora rule optimal?

Obviously will need some assumptions.

Assumptions

Assumption 1 (motivated by u of Weitzman's problem)

- $u(0, x_2, \ldots, x_k) = u(x_2, \ldots, x_k)$; $u(0, \ldots, 0) = 0$;
- u is continuous, nonnegative, symmetric, nondecreasing and submodular in its arguments;

'submodular' means the increase in $u(x)$ obtained by increasing a component of x is nonincreasing as any other component increases.

Assumptions

Assumption 1 (motivated by u of Weitzman's problem)

- $u(0, x_2, \ldots, x_k) = u(x_2, \ldots, x_k)$; $u(0, \ldots, 0) = 0$;
- u is continuous, nonnegative, symmetric, nondecreasing and submodular in its arguments;

'submodular' means the increase in $u(x)$ obtained by increasing a component of x is nonincreasing as any other component increases.

Assumption 2

The benefit of due to opening box j and adding x_j to the set of prizes is independent of the values of already uncovered prizes x_S which are greater than x_j. That is, for $x_j \leq \underline{x}_k < \overline{x}_k$,

$$u(x_S, x_j, \underline{x}_k) - u(x_S, \underline{x}_k) = u(x_S, x_j, \overline{x}_k) - u(x_S, \overline{x}_k).$$

At first sight Assumption 2 appears stronger than we would like. But it is inescapable, as the following lemma makes clear.

Necessity of Assumption 2

Lemma 1 *Suppose the utility u satisfies Assumption 1 and the Pandora rule maximizes expected utility for all distributions F_i and costs c_i. Then u also satisfies Assumption 2.*

Necessity of Assumption 2

Lemma 1 *Suppose the utility u satisfies Assumption 1 and the Pandora rule maximizes expected utility for all distributions F_i and costs c_i. Then u also satisfies Assumption 2.*

Proof. Suppose there were a violation of Assumption 2 of the form, $x_j \leq \underline{x}_k < \overline{x}_k$,

$$u(x_S, x_j, \underline{x}_k) - u(x_S, \underline{x}_k) > u(x_S, x_j, \overline{x}_k) - u(x_S, \overline{x}_k).$$

(By Assumption 1 (submodularity) we can only have $\geq$.)

One shows that if this is true then the Pandora rule cannot be optimal for all $(c_i, F_i,\ i \in N)$, by explicitly constructing an example for which it fails to be optimal. □

More necessity

Lemma 2 *Suppose the utility u satisfies Assumptions 1 and 2. For any $(x_S : i \in S)$, now let $\tilde{x}_\ell$ denote the ℓth greatest element. Then,*

(a) there exist functions $f_\ell : \mathbb{R} \to \mathbb{R}$, $\ell = 1, 2, \ldots$ such that for any x_S we have

$$u(x_S) = \sum_{\ell=1}^{|S|} f_\ell(\tilde{x}_\ell).$$

(b) $f_\ell(0) = 0$.

(c) $f_\ell(x)$ is nondecreasing in x and nonincreasing in ℓ,

(d) $f_\ell(x) - f_{\ell+1}(x)$ nonincreasing in x.

For what utility functions u is a Pandora rule optimal?

Researching this feels like a Pandora box problem!

For what utility functions u is a Pandora rule optimal?

Researching this feels like a Pandora box problem!

Yet even more necessity!

The following is true in special cases, but not yet proved in general.

Conjecture 1 *Suppose the utility u satisfies Assumption 1, and the Pandora rule maximizes expected utility for all distributions F_i and costs c_i. Let $\tilde{x}_1 \geq \tilde{x}_2 \geq \cdots \geq \tilde{x}_{|S|}$ denote the ordered $(x_i : i \in S)$. In particular, $\tilde{x}_1 = \max_{i \in S} x_i$.*

Then necessarily,

$$u(x_S) = u(\tilde{x}_1) - f(\tilde{x}_1) + \sum_{i \in S} f(\tilde{x}_i), \tag{1}$$

where u, f and $u - f$ are all nonnegative, nondecreasing functions.

Yet even more necessity!

The following is true in special cases, but not yet proved in general.

Conjecture 1 *Suppose the utility u satisfies Assumption 1, and the Pandora rule maximizes expected utility for all distributions F_i and costs c_i. Let $\tilde{x}_1 \geq \tilde{x}_2 \geq \cdots \geq \tilde{x}_{|S|}$ denote the ordered $(x_i : i \in S)$. In particular, $\tilde{x}_1 = \max_{i \in S} x_i$.*

Then necessarily,

$$u(x_S) = u(\tilde{x}_1) - f(\tilde{x}_1) + \sum_{i \in S} f(\tilde{x}_i), \qquad (1)$$

where u, f and $u - f$ are all nonnegative, nondecreasing functions.

Is a generalization of Pandora rule evaporating?

Sufficiency

Theorem 1 *Suppose u has the form described as necessary, i.e.*

$$u(x_S) = u(\tilde{x}_1) - f(\tilde{x}_1) + \sum_{i \in S} f(\tilde{x}_i), \tag{2}$$

where u, f and $u - f$ are all nonnegative, nondecreasing functions. Then Pandora rule is optimal for all (c_i, F_i).

Before proof, a digression about bandit processes.

A digression on multi-armed bandits

Google | Gittins "MULTI-ARMED BANDIT"

Web | Images | Maps | Shopping | More | Search tools

About 87,800 results (0.22 seconds)

Scholarly articles for Gittins "MULTI-ARMED BANDIT"
Front Matter - **Gittins** - Cited by 771
Multi-armed bandits and the **Gittins** index - Whittle - Cited by 326
... in a rigged casino: The adversarial **multi-armed bandit** ... - Auer - Cited by 377

Gittins index - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Gittins**_index
If the projects are independent from each other and only one project at a time may evolve, the problem is called **multi-armed bandit** and the **Gittins** index policy is ...
You've visited this page many times. Last visit: 11/3/13

Multi-armed bandit - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Multi-armed_bandit**
In probability theory, the **multi-armed bandit** problem (sometimes called the K- or ... A theorem, the **Gittins** index published first by John C. **Gittins** gives an optimal ...

[PDF] The **Multi-Armed Bandit** Problem: Index Theory Since **Gittins**
www.statslab.cam.ac.uk/~rrw1/talks/gocps.pdf
by R Weber - Related articles
The **Multi-Armed Bandit** Problem: Index Theory Since **Gittins**. Richard Weber. Statistical Laboratory, University of Cambridge. March 2, GOCPS Leipzig 2010 ...

[PDF] A Short Proof of the **Gittins** Index Theorem - MIT
www.mit.edu/~jnt/Papers/J048-94-jnt-**gittins**.pdf
by JN Tsitsiklis - 1994 - Cited by 68 - Related articles
Gittins index theorem for the **multi—armed bandit** problem. The original proof of **Gittins** and Jones [3] and a later proof by **Gittins** [2] relied on an interchange ...

Shop for **Git** **BANDIT" or**

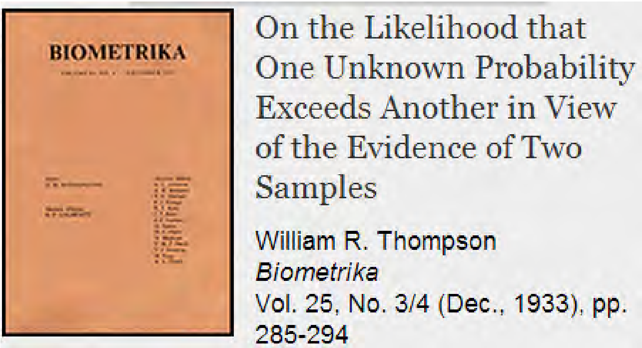
Multiâ–armed Bandit ...
£57.56
eBay

Two-armed bandit

On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples

William R. Thompson
Biometrika
Vol. 25, No. 3/4 (Dec., 1933), pp. 285-294

Two-armed bandit



Robbins, H. (1952). "Some aspects of the sequential design of experiments".

Two-armed bandit

3, 10, 4, 9, 12, 1, ...

5, 6, 2, 15, 2, 7, ...

Two-armed bandit

3, 10, 4, 9, 12, 1, ...

$\longrightarrow$ 5

, 6, 2, 15, 2, 7, ...

Two-armed bandit

3, 10, 4, 9, 12, 1, ...

, , 2, 15, 2, 7, ... $\longrightarrow$ 5, 6

Two-armed bandit

, 10, 4, 9, 12, 1, ...

, , 2, 15, 2, 7, ...

$\longrightarrow$ 5, 6, 3

Two-armed bandit

, , 4, 9, 12, 1, ...

, , 2, 15, 2, 7, ...

$\longrightarrow$ 5, 6, 3, 10,

Two-armed bandit

, , , 9, 12, 1, ...

, , 2, 15, 2, 7, ...

$\longrightarrow$ 5, 6, 3, 10, 4

Two-armed bandit

, , , , 12, 1, ...

, , 2, 15, 2, 7, ... $\longrightarrow$ 5, 6, 3, 10, 4, 9

Two-armed bandit

, , , , , 1, ...

, , 2, 15, 2, 7, ...

$\longrightarrow$ 5, 6, 3, 10, 4, 9, 12

Two-armed bandit

, , , , , 1, ...

, , , 15, 2, 7, ...

$\longrightarrow$ 5, 6, 3, 10, 4, 9, 12, 2

Two-armed bandit

$, \quad , \quad , \quad , \quad , 1, \ldots$

$, \quad , \quad , \quad , 2, 7, \ldots$

$\longrightarrow$ 5, 6, 3, 10, 4, 9, 12, 2, 15

Two-armed bandit

, , , , , 1, ...

, , , , 2, 7, ... $\longrightarrow$ 5, 6, 3, 10, 4, 9, 12, 2, 15

Reward = $5 + 6\,\beta + 3\,\beta^2 + 10\,\beta^3 + \cdots$

$0 < \beta < 1.$

Two-armed bandit

$$, \quad , \quad , \quad , \quad , 1, \ldots$$

$$, \quad , \quad , \quad , 2, 7, \ldots$$

$$\longrightarrow 5, 6, 3, 10, 4, 9, 12, 2, 15$$

$$\text{Reward} = 5 + 6\,\beta + 3\,\beta^2 + 10\,\beta^3 + \cdots$$

$0 < \beta < 1$. Of course, in practice we must choose which arms to pull without knowing the future sequences of rewards.

Each of the two arms is a **bandit process**.

Bandit processes

A bandit process is a special type of Markov Decision Process in which there are just two possible actions:

- $u = 1$ (**continue**)

  produces reward $r(x_t)$ and the state changes, to x_{t+1}, according to Markov dynamics $P_i(x_t, x_{t+1})$.

- $u = 0$ (**freeze**)

  produces no reward and the state does not change (hence the term 'freeze').

Bandit processes

A bandit process is a special type of Markov Decision Process in which there are just two possible actions:

- $u = 1$ (**continue**)

  produces reward $r(x_t)$ and the state changes, to x_{t+1}, according to Markov dynamics $P_i(x_t, x_{t+1})$.

- $u = 0$ (**freeze**)

  produces no reward and the state does not change (hence the term 'freeze').

A **simple family of alternative bandit processes** (SFABP) is a collection of N such bandit processes, in known states $x_1(t), \ldots, x_N(t)$.

SFABP

At each time, $t \in \{0, 1, 2, \ldots\}$,

- One bandit process is to be activated (pulled/**continued**)

  If arm i activated then it changes state:

$$x \to y \quad \text{with probability } P_i(x, y)$$

  and produces reward $r_i(x_i(t))$.

SFABP

At each time, $t \in \{0, 1, 2, \ldots\}$,

- One bandit process is to be activated (pulled/**continued**)

  If arm i activated then it changes state:

  $$x \to y \quad \text{with probability } P_i(x, y)$$

  and produces reward $r_i(x_i(t))$.
- All other bandit processes remain passive (not pulled/**frozen**).

SFABP

At each time, $t \in \{0, 1, 2, \ldots\}$,

- One bandit process is to be activated (pulled/**continued**)

  If arm i activated then it changes state:

$$x \to y \quad \text{with probability } P_i(x, y)$$

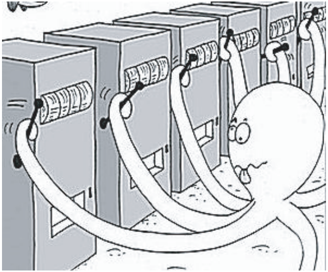
  and produces reward $r_i(x_i(t))$.
- All other bandit processes remain passive (not pulled/**frozen**).

Objective: maximize the expected total β-discounted reward

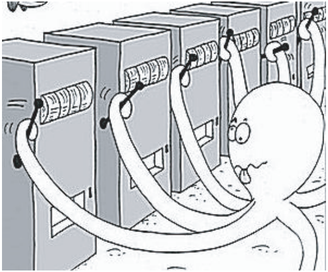
$$E\left[\sum_{t=0}^{\infty} r_{i_t}(x_{i_t}(t))\, \beta^t\right],$$

where i_t is the arm pulled at time t, $(0 < \beta < 1)$.

Dynamic effort allocation

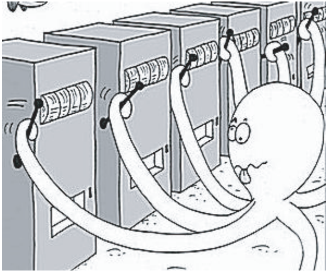


Dynamic effort allocation



- **Job Scheduling**: in what order should I work on the tasks in my in-tray?
- **Research projects**: how should I allocate my research time amongst my favorite open problems so as to maximize the value of my completed research?

Dynamic effort allocation



- **Searching for information**: shall I spend more time browsing the web, or go to the library, or ask a friend?
- **Dating strategy**: should I contact a new prospect, or try another date with someone I have dated before?

Dynamic programming for bandit processes

The dynamic programming equation is

$$F(x_1, \dots, x_N)$$
$$= \max_i \Big\{ r_i(x_i) + \beta \sum_y P_i(x_i, y) F(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_N) \Big\}$$

Dynamic programming for bandit processes

The dynamic programming equation is

$$F(x_1, \ldots, x_N)$$

$$= \max_i \Big\{ r_i(x_i) + \beta \sum_y P_i(x_i, y) F(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_N) \Big\}$$

If bandit i moves on a state space of size k_i, then $(x_1, \ldots, x_N)$ moves on a state space of size $\prod_i k_i$ (exponential in N).

Gittins index theorem

Theorem [Gittins, '74, '79, '89]

The expected discounted reward obtained from a simple family of alternative bandit processes is maximized by always continuing the bandit having greatest Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\Big[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \,\Big|\, x_i(0) = x_i \Big]}{E\Big[\sum_{t=0}^{\tau-1} \beta^t \,\Big|\, x_i(0) = x_i \Big]}.$$

where τ is a (past-measurable) stopping-time.

Gittins index theorem

Theorem [Gittins, '74, '79, '89]

The expected discounted reward obtained from a simple family of alternative bandit processes is maximized by always continuing the bandit having greatest Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \;\middle|\; x_i(0) = x_i \right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \;\middle|\; x_i(0) = x_i \right]}.$$

where τ is a (past-measurable) stopping-time.

$G_i(x_i)$ is called the **Gittins index**.

Gittins index theorem

Theorem [Gittins, '74, '79, '89]

The expected discounted reward obtained from a simple family of alternative bandit processes is maximized by always continuing the bandit having greatest Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \,\middle|\, x_i(0) = x_i\right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \,\middle|\, x_i(0) = x_i\right]}.$$

where τ is a (past-measurable) stopping-time.

$G_i(x_i)$ is called the **Gittins index**.

Gittins and Jones (1974). A dynamic allocation index for the sequential design of experiments. In Gani, J., editor, Progress in Statistics, pages 241–66. North-Holland, Amsterdam, NL. Read at the 1972 European Meeting of Statisticians, Budapest.

Gittins index for bandits

Like Weitzman's reservation value the Gittins index can be defined by calibration. Suppose there is just one bandit process, B_i.

What bandit process, B_0, producing constant reward G_i per 'pull', would if offered as an alternative to B_i, make us indifferent as to which of these two bandits to continue next?

i.e. as a function of the current state $x_i(0) = x$,

$$(1 + \beta + \beta^2 + \cdots)G_i(x)$$
$$= \sup_{\tau \geq 1} E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t + (\beta^\tau + \beta^{\tau+1} + \cdots)G_i(x) \,\middle|\, x_i(0) = x \right]$$

where supremum is over time τ of switching from B_0 to B_i.

Gittins index for bandits

Like Weitzman's reservation value the Gittins index can be defined by calibration. Suppose there is just one bandit process, B_i.

What bandit process, B_0, producing constant reward G_i per 'pull', would if offered as an alternative to B_i, make us indifferent as to which of these two bandits to continue next?

i.e. as a function of the current state $x_i(0) = x$,

$$(1 + \beta + \beta^2 + \cdots)G_i(x)$$

$$= \sup_{\tau \geq 1} E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t + (\beta^\tau + \beta^{\tau+1} + \cdots)G_i(x) \,\middle|\, x_i(0) = x \right]$$

where supremum is over time τ of switching from B_0 to B_i. Equivalently,

$$G_i(x) = \sup_{\tau > 0} \frac{E\left[\sum_{t=0}^{\tau-1} \beta^t r_i(x_i(t) \,\middle|\, x_i(0) = x \right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \,\middle|\, x_i(0) = x \right]}.$$

Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \;\middle|\; x_i(0) = x_i \right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \;\middle|\; x_i(0) = x_i \right]}$$

Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\Big[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \ \Big|\ x_i(0) = x_i \Big]}{E\Big[\sum_{t=0}^{\tau-1} \beta^t \ \Big|\ x_i(0) = x_i \Big]}$$

Discounted reward up to τ.

Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \;\middle|\; x_i(0) = x_i \right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \;\middle|\; x_i(0) = x_i \right]}$$

Discounted reward up to τ.

Discounted time up to τ.

Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \;\middle|\; x_i(0) = x_i\right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \;\middle|\; x_i(0) = x_i\right]}$$

Discounted reward up to τ.

Discounted time up to τ.

Note the role of the **stopping time** τ.
Stopping times are times recognisable when they occur.

Gittins index

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E\left[\sum_{t=0}^{\tau-1} r_i(x_i(t))\beta^t \;\middle|\; x_i(0) = x_i \right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \;\middle|\; x_i(0) = x_i \right]}$$

Discounted reward up to τ.

Discounted time up to τ.

Note the role of the **stopping time** τ.
Stopping times are times recognisable when they occur.
How do you make perfect toast?

There is a rule for timing toast,
One never has to guess,
Just wait until it starts to smoke,
then 7 seconds less. (David Kendall)

A short history of the index theorem

A short history of the index theorem

Many applications to clinical trials, job scheduling, search, etc.

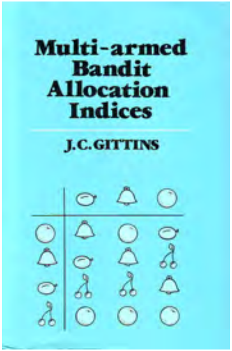
A short history of the index theorem

Exploration vs Exploitation

"Bandit problems embody in essential form a conflict evident in all human action: information versus immediate payoff." (Whittle)

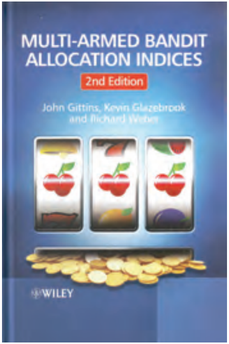
Many applications to clinical trials, job scheduling, search, etc.

A short history of the index theorem



Many applications to clinical trials, job scheduling, search, etc.

A short history of the index theorem



Many applications to clinical trials, job scheduling, search, etc.

A short history of the index theorem

Many applications to clinical trials, job scheduling, search, etc.

Gittins index theorem is surprising!

Peter Whittle tells the story:

“A colleague of high repute asked an equally well-known colleague:

— *What would you say if you were told that the multi-armed bandit problem had been solved?’*

Gittins index theorem is surprising!

Peter Whittle tells the story:

"A colleague of high repute asked an equally well-known colleague:

— *What would you say if you were told that the multi-armed bandit problem had been solved?'*

— *Sir, the multi-armed bandit problem is not of such a nature that it can be solved.'*

Proofs of the Index Theorem

Since Gittins (1974, 1979), many researchers have reproved, remodelled and resituated the index theorem.

Beale (1979)

Karatzas (1984)

Varaiya, Walrand, Buyukkoc (1985)

Chen, Katehakis (1986)

Kallenberg (1986)

Katehakis, Veinott (1986)

Eplett (1986)

Kertz (1986)

Tsitsiklis (1986)

Mandelbaum (1986, 1987)

Lai, Ying (1988)

Whittle (1988)

Weber (1992)

El Karoui, Karatzas (1993)

Ishikida and Varaiya (1994)

Tsitsiklis (1994)

Bertsimas, Niño-Mora (1996)

Glazebrook, Garbe (1996)

Kaspi, Mandelbaum (1998)

Bäuerle, Stidham (2001)

Dimitriu, Tetali, Winkler (2003)

Golf with N balls

Dumitriu, Tetali and Winkler, (2003). On playing golf with two balls.

N balls are strewn about a golf course at locations $x_1, \ldots, x_N$.

Golf with N balls

N balls are strewn about a golf course at locations $x_1, \ldots, x_N$.

Hitting a ball i, that is in location x_i, costs $c(x_i)$,

$$x_i \to y \quad \text{with probability } P(x_i, y)$$

Ball goes in the hole with probability $P(x_i, 0)$.

Objective

Minimize the expected total cost incurred until sinking a first ball.

Golf with N balls

N balls are strewn about a golf course at locations $x_1, \ldots, x_N$.

Hitting a ball i, that is in location x_i, costs $c(x_i)$,

$$x_i \to y \quad \text{with probability } P(x_i, y)$$

Ball goes in the hole with probability $P(x_i, 0)$.

Objective

Minimize the expected total cost incurred until sinking a first ball.

Answer

When ball i is in location x_i it has an index $\gamma_i(x_i)$.

Play the ball of smallest index, until a ball goes in the whole.

Gittins index theorem for golf with N balls

Golf with one ball

Consider golf with one ball, initially in location x_i.

Offer golfer a prize λ, obtained when ball goes in the hole (state 0).

Gittins index theorem for golf with N balls

Golf with one ball

Consider golf with one ball, initially in location x_i.

Offer golfer a prize λ, obtained when ball goes in the hole (state 0).

We might ask, *what is the least λ for which it is optimal for him to take at least one more stroke* — allowing him the option to retire at any point thereafter?

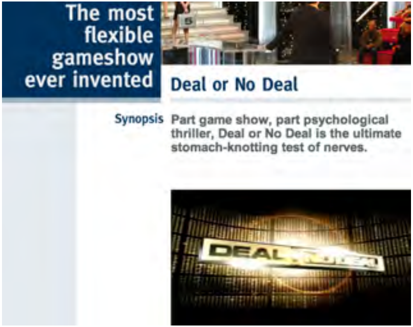
Gittins index theorem for golf with N balls

Golf with one ball

Consider golf with one ball, initially in location x_i.

Offer golfer a prize λ, obtained when ball goes in the hole (state 0).

We might ask, *what is the least λ for which it is optimal for him to take at least one more stroke* — allowing him the option to retire at any point thereafter?



Gittins index theorem for golf with N balls

Golf with one ball

Consider golf with one ball, initially in location x_i.

Offer golfer a prize λ, obtained when ball goes in the hole (state 0).

We might ask, *what is the least λ for which it is optimal for him to take at least one more stroke* — allowing him the option to retire at any point thereafter?

$$\lambda_i(x_i) = \inf \left\{ \lambda : 0 \leq \sup_{\tau \geq 1} E \left[\lambda 1_{\{x_i(\tau)=0\}} - \sum_{t=0}^{\tau-1} c_i(x_i(t) \,\middle|\, x_i(0) = x_i \right] \right\}.$$

Call $\lambda_i(x_i)$ the **fair prize**, (or Gittins index).

How to play golf with one ball and an increasing fair prize

Having been offered a fair prize the golfer will play until the ball

- goes in the hole, or
- reaches a state $x_i(t)$ from which the offered prize is no longer great enough to tempt him to play further.

How to play golf with one ball and an increasing fair prize

Having been offered a fair prize the golfer will play until the ball

- goes in the hole, or
- reaches a state $x_i(t)$ from which the offered prize is no longer great enough to tempt him to play further.

  If the latter occurs, let us increase the prize to $\lambda_i(x_i(t))$.

It becomes the '**prevailing prize**' at t, i.e.
$\gamma_i(t) = \max_{0 \leq s \leq t} \lambda_i(x_i(s))$, which is nondecreasing in t.

How to play golf with one ball and an increasing fair prize

Having been offered a fair prize the golfer will play until the ball

- goes in the hole, or
- reaches a state $x_i(t)$ from which the offered prize is no longer great enough to tempt him to play further.

  If the latter occurs, let us increase the prize to $\lambda_i(x_i(t))$.

It becomes the '**prevailing prize**' at t, i.e.
$\gamma_i(t) = \max_{0 \leq s \leq t} \lambda_i(x_i(s))$, which is nondecreasing in t.

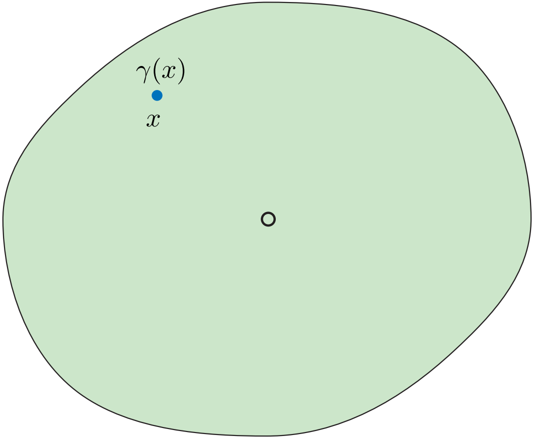
Now the golfer need never retire and can keep playing until the ball goes in the hole, say at time τ.

But his expected profit is just 0.

$$E\left[\gamma_i(x_i(\tau-1)) - \sum_{t=0}^{\tau-1} c_i(x_i(t) \,\middle|\, x_i(0) = x_i\right] = 0.$$

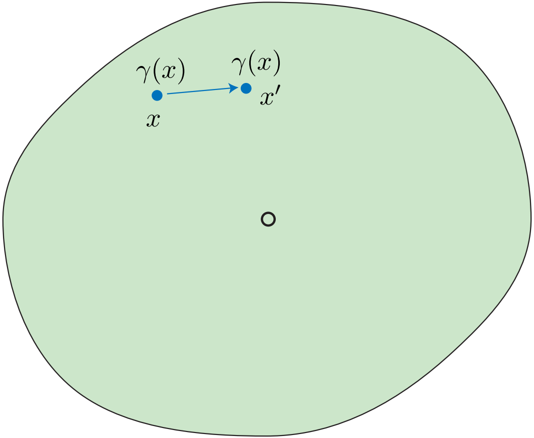
Golf with 1 ball

$\gamma(x) = 3.0$



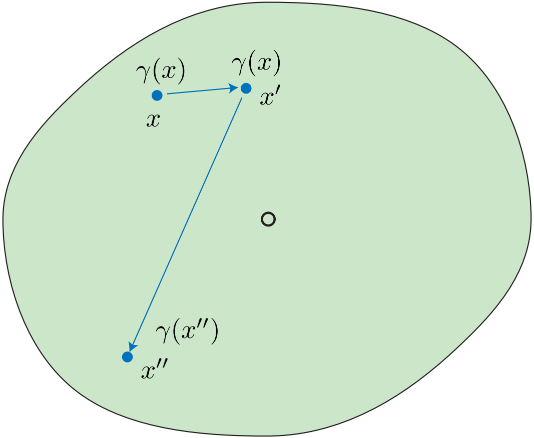
Golf with 1 ball

$\gamma(x) = 3.0$, $\gamma(x') = 2.5$



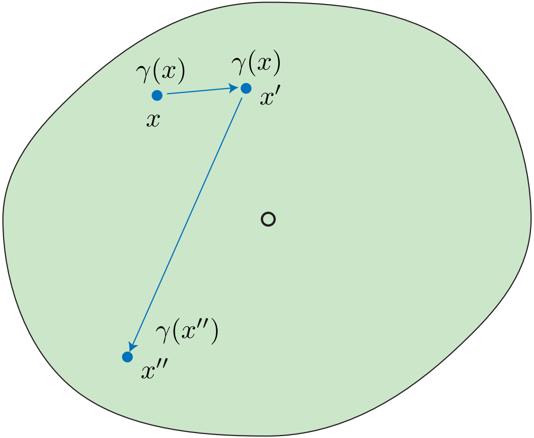
Golf with 1 ball

$\gamma(x) = 3.0$, $\gamma(x') = 2.5$, $\gamma(x'') = 4.0$



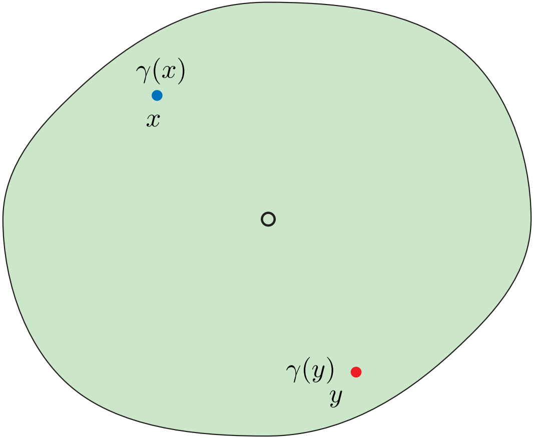
Golf with 1 ball

$\gamma(x) = 3.0$, $\gamma(x') = 2.5$, $\gamma(x'') = 4.0$
Prevailing prize sequence is 3.0, 3.0, 4.0, $\ldots$



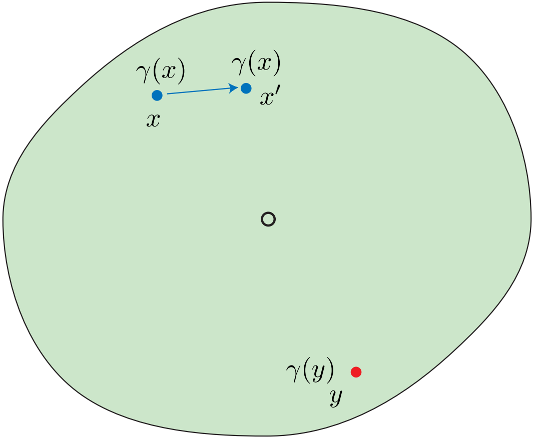
Golf with 2 balls

$\gamma(x) = \overline{3.0}$
$\gamma(y) = \overline{3.2}$



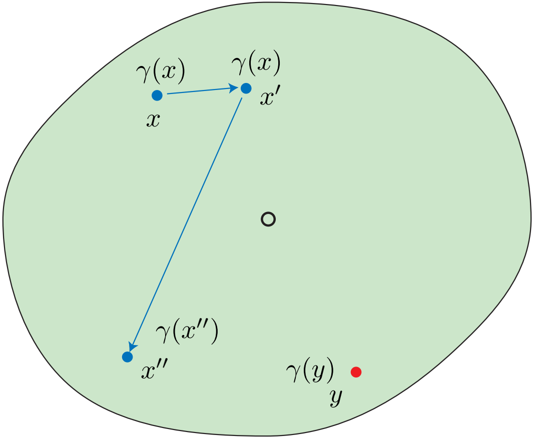
Golf with 2 balls

$\gamma(x) = \overline{3.0}$, $\gamma(x') = 2.5$
$\gamma(y) = \overline{3.2}$



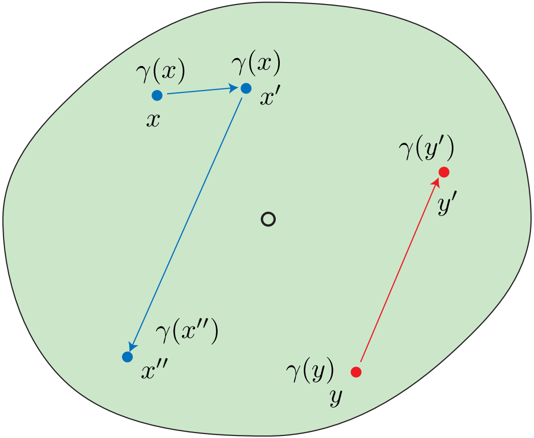
Golf with 2 balls

$\gamma(x) = 3.0$, $\gamma(x') = 2.5$, $\gamma(x'') = \overline{4.0}$
$\color{red}{\gamma(y) = \overline{3.2}}$



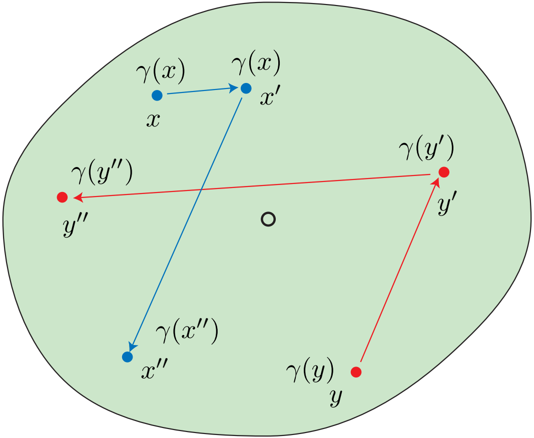
Golf with 2 balls

$\gamma(x) = 3.0$, $\gamma(x') = 2.5$, $\gamma(x'') = \overline{4.0}$
$\gamma(y) = 3.2$, $\gamma(y') = \overline{3.5}$



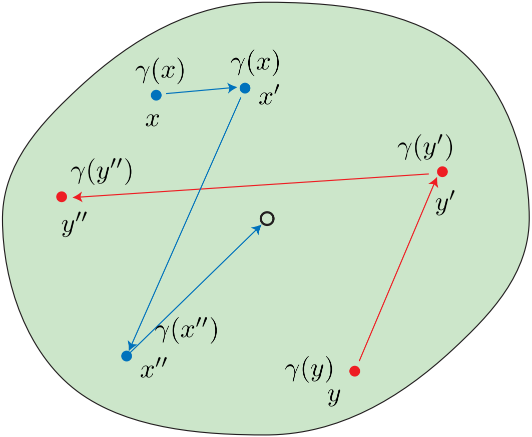
Golf with 2 balls

$\gamma(x) = 3.0$, $\gamma(x') = 2.5$, $\gamma(x'') = \overline{4.0}$
$\gamma(y) = 3.2$, $\gamma(y') = 3.5$, $\gamma(y'') = \overline{4.2}$



Golf with 2 balls

$\gamma(x) = 3.0$, $\gamma(x') = 2.5$, $\gamma(x'') = \overline{4.0}$
$\gamma(y) = 3.2$, $\gamma(y') = 3.5$, $\gamma(y'') = \overline{4.2}$



Index theorem for golf with N balls

Suppose the golfer keeps playing until a ball goes in the hole.

His prize is the prevailing prize of the ball he sinks.

Index theorem for golf with N balls

Suppose the golfer keeps playing until a ball goes in the hole.

His prize is the prevailing prize of the ball he sinks.

Prevailing prizes are defined in such a way that the golfer cannot make a strictly positive profit, and so for any policy σ,

$$E_\sigma(\text{cost incurred}) \geq E_\sigma(\text{prize eventually won}) \tag{1}$$

Index theorem for golf with N balls

Suppose the golfer keeps playing until a ball goes in the hole.

His prize is the prevailing prize of the ball he sinks.

Prevailing prizes are defined in such a way that the golfer cannot make a strictly positive profit, and so for any policy σ,

$$E_\sigma(\text{cost incurred}) \geq E_\sigma(\text{prize eventually won}) \qquad (1)$$

Let π be the policy: *always play the ball with least prevailing prize.*

Because each ball's sequence of prevailing prizes is nondecreasing.

$$E_\sigma(\text{prize eventually won}) \geq E_\pi(\text{prize eventually won}) \qquad (2)$$

Index theorem for golf with N balls

Suppose the golfer keeps playing until a ball goes in the hole.

His prize is the prevailing prize of the ball he sinks.

Prevailing prizes are defined in such a way that the golfer cannot make a strictly positive profit, and so for any policy σ,

$$E_\sigma(\text{cost incurred}) \geq E_\sigma(\text{prize eventually won}) \tag{1}$$

Let π be the policy: *always play the ball with least prevailing prize.*

Because each ball's sequence of prevailing prizes is nondecreasing.

$$E_\sigma(\text{prize eventually won}) \geq E_\pi(\text{prize eventually won}) \tag{2}$$

But the golfer breaks even under π.

$$E_\pi(\text{prize eventually won}) = E_\pi(\text{cost incurred}) \tag{3}$$

Generalizing Weitzman's Pandora problem

Gittins index theorem and Weitzman's problem

Theorem (Gittins index theorem, 1972) *The problem posed by a family of alternative bandit processes, is solved by always continuing the bandit process having the greatest Gittins index.*

Compare this to the solution to the Weitzman's problem which is

Theorem (Weitzman's Pandora rule, 1979). *Pandora's problem is solved by always opening the unopened box with greatest reservation value, until all reservations values are less than the greatest prize that has been found.*

Pandora plays golf

Learn how to play golf with more than one ball.

- You can solve Generalized Pandora's boxes problem.

Pandora plays golf

Learn how to play golf with more than one ball.

- You can solve Generalized Pandora's boxes problem.

- Pandora's box i is now ball i, starting in state 1, say.

Pandora plays golf

Learn how to play golf with more than one ball.

- You can solve Generalized Pandora's boxes problem.

- Pandora's <u>box i</u> is now <u>ball i</u>, starting in state 1, say.
- First time ball i is hit, cost c_i is incurred, and ball lands at location $(x_i, 1)$ (where x_i sampled from F_i).

Pandora plays golf

Learn how to play golf with more than one ball.

- You can solve Generalized Pandora's boxes problem.

- Pandora's <u>box i</u> is now <u>ball i</u>, starting in state 1, say.
- First time ball i is hit, cost c_i is incurred, and ball lands at location $(x_i, 1)$ (where x_i sampled from F_i).
- Second time ball i is hit, (from current state $(x_i, 1)$), cost $-f(x_i)$ is incurred, the ball goes to state $(x_i, 2)$.

Pandora plays golf

Learn how to play golf with more than one ball.

- You can solve Generalized Pandora's boxes problem.

- Pandora's box i is now ball i, starting in state 1, say.
- First time ball i is hit, cost c_i is incurred, and ball lands at location $(x_i, 1)$ (where x_i sampled from F_i).
- Second time ball i is hit, (from current state $(x_i, 1)$), cost $-f(x_i)$ is incurred, the ball goes to state $(x_i, 2)$.
- Third time ball i is hit, (from current state $(x_i, 2)$), cost $f(x_i) - u(x_i)$ is incurred, it goes in hole, 0, and game ends.

Pandora plays golf

Learn how to play golf with more than one ball.

- You can solve Generalized Pandora's boxes problem.

- Pandora's <u>box i</u> is now <u>ball i</u>, starting in state 1, say.
- First time ball i is hit, cost c_i is incurred, and ball lands at location $(x_i, 1)$ (where x_i sampled from F_i).
- Second time ball i is hit, (from current state $(x_i, 1)$), cost $-f(x_i)$ is incurred, the ball goes to state $(x_i, 2)$.
- Third time ball i is hit, (from current state $(x_i, 2)$), cost $f(x_i) - u(x_i)$ is incurred, it goes in hole, 0, and game ends.

The following two problems are equilvalent.

- Minimize the expected cost of putting a ball in the hole
- Maximize the expected value of Pandora's greatest discovered prize, net of costs of opening boxes.

Gittins $\implies$ **Weitzman**, (is mentioned by Chade and Smith (2006))

Gittins index for generalized Pandora

The Gittins index of ball i is its prevailing prize, γ_i, which in state 0 is the least γ such that

$$0 \leq -c_i + Ef(x_i) + E\max\{0, u(x_i) - f(x_i) + \gamma\}.$$

The generalized reservation value is the least x_i^* such that

$$u(x_S, x_i^*) \geq -c_i + E[u(x_S, x_i, x_i^*)\}]$$

Easy to check $\gamma_i = x_i^*$, and so prescription of Gittins index theorem is identical to Pandora rule with generalized reservation values.

Is Gittins index the whole story?

We seem to have come to the (disappointing?) conclusion that a Pandora rule is optimal iff

$$u(x_S) = u(x_1) - f(x_1) + \sum_{i \in S} f(x_i),$$

and the 'if' part follows from an application of the Gittins index theorem.

So is there any more to say?

A problem which cannot be solved by Gittins index theorem

- Pandora has n boxes.
- Box i is either empty (w.p. q_i) or has €1 (w.p. p_i)
- Costs c_i to take content of box i.
- Wish to maximize expected value of

$$\psi\left(\sum_{i \in S} x_i\right) - \sum_{i \in S} c_i$$

where ψ is concave increasing of total wealth.

- Having found €k the reservation value of unopened box i is least y such that

$$\psi(k+y) \geq -c_i + p_i \psi(k+y+1) + q_i \psi(k+y)$$

least y such that

$$c_i/p_i \geq \psi(k+y+1) - \psi(k+y)$$

and so the x_i^* are ordered in the same way as c_i/p_i.
Suppose $c_1/p_1 \leq c_2/p_2 \leq \cdots \leq c_n/p_n$.

Pandora rule: Open the boxes in the order $1, 2, \ldots, n$, stopping when we are about to open some box j, have accumulated €k, $k \leq j-1$, and

$$c_j/p_j > \psi(k+1) - \psi(k).$$

While it would also be possible to guess this answer, and establish optimality of the Pandora rule by a fairly short proof tailored to this problem and using induction on n, the sufficient conditions provided by Theorem 2 are quick to check.

A sufficient condition for Pandora rule optimality

History-Independence of the Ordering of Reservation Values (ORD): The ordering of reservation values x_k^* of the covered variables is independent of both the number of variables that have already been uncovered and their realizations. That is, for any S, x_S, and $k, j \notin S$,

$$x_k^*(x_S) \geq x_j^*(x_S) \quad \iff \quad x_k^*(\varnothing) \geq x_j^*(\varnothing)$$

This is a joint property of the utility function u, costs c_i and distributions F_i.

A sufficient condition for Pandora rule optimality

History-Independence of the Ordering of Reservation Values (ORD): The ordering of reservation values x_k^* of the covered variables is independent of both the number of variables that have already been uncovered and their realizations. That is, for any S, x_S, and $k, j \notin S$,

$$x_k^*(x_S) \geq x_j^*(x_S) \quad \iff \quad x_k^*(\varnothing) \geq x_j^*(\varnothing)$$

This is a joint property of the utility function u, costs c_i and distributions F_i.

Theorem 2 *If the utility function u satisfies Assumptions 1, 2, and ORD then the Pandora rule maximizes expected utility.*

Back to the open conjecture

Conjecture 1 *Suppose the utility u satisfies Assumption 1, and the Pandora rule maximizes expected utility for all distributions F_i and costs c_i. Let $x_1 \geq x_2 \geq \cdots \geq x_{|S|}$ denote the ordered $(x_i : i \in S)$. In particular, $x_1 = \max_{i \in S} x_i$.*

Then necessarily,

$$u(x_S) = u(x_1) - f(x_1) + \sum_{i \in S} f(x_i), \tag{3}$$

where u, f and $u - f$ are all nonnegative, nondecreasing functions.

Conjecture has been proved, ... but not if we require all reservation values to be finite.

A special case

Lemma 2 *Suppose the utility* u *satisfies Assumptions 1 and 2. Let* $\tilde{x}_1 \geq \tilde{x}_2 \geq \cdots \geq \tilde{x}_{|S|}$ *denote the ordered* $(x_i : i \in S)$*. Then,*

$$u(x_S) = \sum_{\ell=1}^{|S|} f_\ell(\tilde{x}_\ell).$$

where $f_1 \geq f_2 \geq \cdots \geq f_{|S|}$.

A special case

Lemma 2 *Suppose the utility u satisfies Assumptions 1 and 2. Let $\tilde{x}_1 \geq \tilde{x}_2 \geq \cdots \geq \tilde{x}_{|S|}$ denote the ordered $(x_i : i \in S)$. Then,*

$$u(x_S) = \sum_{\ell=1}^{|S|} f_\ell(\tilde{x}_\ell).$$

where $f_1 \geq f_2 \geq \cdots \geq f_{|S|}$.

Theorem 3 *Suppose conditions of Lemma 2 and consider the special case $f_\ell(x) = w_\ell x$, so the utility u is*

$$u(x_S) = \sum_{i=1}^{|S|} w_i \tilde{x}_i$$

where $w_1 \geq w_2 \geq w_3 \geq \cdots$ are given. If the Pandora rule is to be optimal for all (c_i, F_i) then necessarily, $w_2 = w_3 = w_4 = \cdots$.

Proof of Theorem 3

Proof is interesting.

- Start with 3 boxes and construct distributions that force us to conclude that if the Pandora rule is optimal then $w_2 = w_3$. One box is of 'type A' and two boxes are of 'type B'.

  Type A box has prize taking values 0,1,2 with probabilities α, β, γ respectively, where $(\alpha, \beta, \gamma) = (\frac{1}{2}, \frac{3}{8}, \frac{1}{8})$.
  Type B variable is similar, but $(\alpha, \beta, \gamma) = (0, \frac{1}{2}, \frac{1}{2})$.

- Taking $c_A = 3w_2/8$; $c_B = 3w_2/2$ can show that boxes have equal reservation values $x_A^* = x_B^* = 2$.

- Is it really optimal to start by opening either box?

Mathematica program

Can discover the answer by dynamic programming.

```
{p0,p1,p2}={1/2,3/8,1/8};
{q0,q1,q2}={0,1/2,1/2};
y=2;
Solve[y w1 == -cA + p0 w1 y + p1 (w1 y + w2 1) + p2 (w1 2 + w2 y), cA];
cA=cA /.%[[1]];
Solve[y w1 == -cB + q0 w1 y + q1 (w1 y + w2 1) + q2 (w1 2 + w2 y), cB];
cB=cB /.%[[1]];
cA=3 w2/8; cB=3 w2/2;

v[a_, b_, c_] := Sort[{a, b, c}].{w3, w2, w1}
vA[a_, b_]:= Max[v[a, b, 0], -cA + p0 v[a, b, 0] + p1 v[a, b, 1] + p2 v[a, b, 2]]
vB[a_, b_]:= Max[v[a, b, 0], -cB + q0 v[a, b, 0] + q1 v[a, b, 1] + q2 v[a, b, 2]]
vBB[a_]    := Max[v[a, 0, 0], -c2B+ q0 vB[a, 0] + q1 vB[a, 1] + q2 vB[a, 2]]
vAB[a_]    := Max[v[a, 0, 0], -cA + p0 vB[a, 0] + pi vB[a, 1] + p2 vB[a, 2],
                              -cB + q0 vA[a, 0] + q1 vA[a, 1] + q2 vA[a, 2]]
vABB       := Max[0, -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2],
                    -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]]
```

Mathematica as proof engine

Can test by input numeric values of w_1, w_2, w_3, and can also Mathematica as a 'proof engine'.

```
vABB      := Max[0, -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2],
                     -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]]

Assuming[{w1>w2>w3},  0 < -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2]
           < -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]) // Simplify]
Out[1]=
     True

Assuming[{w1>w2==w3}, 0 < -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2]
          == -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]) // Simplify]
Out[2]=
     True
```

Mathematica as proof engine

Can test by input numeric values of w_1, w_2, w_3, and can also Mathematica as a 'proof engine'.

```
vABB      := Max[0, -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2],
                   -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]]

Assuming[{w1>w2>w3},  0 < -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2]
           < -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]) // Simplify]
Out[1]=
     True

Assuming[{w1>w2==w3}, 0 < -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2]
          == -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]) // Simplify]
Out[2]=
     True
```

Out[1] shows that if $w_1 > w_2$ then Pandora rule is not optimal, as it is better to start opening a type B box than the one of type A.

So if the Pandora rule is optimal then necessarily $w_2 = w_3$. $\square$

Mathematica as proof engine

Can test by input numeric values of w_1, w_2, w_3, and can also Mathematica as a 'proof engine'.

```
vABB      := Max[0, -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2],
                   -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]]

Assuming[{w1>w2>w3},  0 < -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2]
           < -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]) // Simplify]
Out[1]=
     True

Assuming[{w1>w2==w3}, 0 < -cA + p0 vBB[a, 0] + pi vBB[a, 1] + p2 vBB[a, 2]
          == -cB + q0 vAB[a, 0] + q1 vAB[a, 1] + q2 vAB[a, 2]) // Simplify]
Out[2]=
     True
```

Presumably, the manipulations that *Mathematica* makes to prove the truth of these inequalities can be replicated 'by hand'.
Is that important?

Digression: an open problem

Consider the same set up of Weitzman's problem.

But now 'offers' do <u>not</u> remain open.

On opening a box Pandora must immediately take the prize it reveals, or reject it with no opportunity to recall.

- Easily 'solved' in the special case that $F_i = F$, i.e. the distribution of the prize value in all boxes is the same.
- But if the F_i differ then even $c_i = 0$ is hard.

  Is it better to open first a box that is likely to contain a large prize, or small prize, or a highly variable prize?

‘Four Rooms’ on Channel 4 TV

‘*People who believe they have a valuable artifact get a chance to sell it to some of the country’s leading dealers. But, once they turn down an offer, there’s no going back...*’

Fred Astaire’s suitcases, a Dali sculpture, a dress made from car parts and a rare Patek watch are amongst the collectibles members of the public are hoping to exchange for life-changing sums.

A More General Pandora's Rule?

A More General Pandora's Rule?

Reading list

[1] Chade, H. and Smith, L., Simultaneous search. Econometrica, 74(5):1293–1307, 2006.

[2] Dumitriu, I., Tetali, P., and Winkler, P., On playing golf with two balls. SIAM J. Discret. Math., 16(4):604–615, 2003

[3] Gittins, J. C., Glazebrook K. D., Weber, R. R., Multiarmed Bandit Allocation Indices, (2nd editon), Wiley, 2011.

[4] Varian, H. R., Economics and Search, SIGIR Forum, Volume 33(1), 1–5, 1999.

[5] Weber, R. R. and Wojciech, O., A more general Pandoras rule?, working paper, 2013.

[6] Weitzman, M. L., Optimal search for the best alternative, Econometrica, 47(3), 1979.

Appendix

Proof of Theorem 2

For ease of explanation we prove this for the special case of Weitzman's problem. The application to more general u is essentially the same, mainly a matter of notation.

Consider first the case of just two boxes, when the reservation value of box 1 is less than that of box 2, i.e. $x_1^* < x_2^*$.

- The optimal strategy, π_1, contingent on box 1 being opened first and revealing prize x_1 is to open box 2 iff $x_1 \leq x_2^*$.
- The optimal strategy contingent on box 2 being opened first and revealing prize x_2 is to open box 1 iff $x_2 \leq x_1^*$.

  A suboptimal strategy, π_2, contingent on box 2 being opened first and revealing prize x_2 is to open box 1 iff $x_2 \leq x_2^*$.
- We show that π_2 is at least as good as π_1.

Table below shows payoffs of π_1 and π_2 contingent on realizations of x_1 and x_2.

Divided into four cells, depending on whether each x_i is below or above the cutoff x_2^*.

Within each cell

- upper rows are payoffs for strategy π_1
- lower rows are payoffs for strategy π_2

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	$x_2 - c_1 - c_2$	$x_1 - c_1$
	$x_2 - c_2$	$x_2 - c_2$
$x_2 < x_2^*$	$\max\{x_1, x_2\} - c_1 - c_2$	$x_1 - c_1$
	$\max\{x_1, x_2\} - c_1 - c_2$	$x_1 - c_1 - c_2$

Table below shows payoffs of π_1 and π_2 contingent on realizations of x_1 and x_2.

Divided into four cells, depending on whether each x_i is below or above the cutoff x_2^*.

Within each cell

- upper rows are payoffs for strategy π_1
- lower rows are payoffs for strategy π_2
- we now cancel things that are the same for π_1 and π_2.

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	$x_2 - c_1 - c_2$ $x_2 - c_2$	$x_1 - c_1$ $x_2 - c_2$
$x_2 < x_2^*$	$\max\{x_1, x_2\} - c_1 - c_2$ $\max\{x_1, x_2\} - c_1 - c_2$	$x_1 - c_1$ $x_1 - c_1 - c_2$

Table below shows payoffs of π_1 and π_2 contingent on realizations of x_1 and x_2.

Divided into four cells, depending on whether each x_i is below or above the cutoff x_2^*.

Within each cell

- upper rows are payoffs for strategy π_1
- lower rows are payoffs for strategy π_2
- we now cancel things that are the same for π_1 and π_2.

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	$x_2 - c_1 - c_2$ $x_2 - c_2$	$x_1 - c_1$ $x_2 - c_2$
$x_2 < x_2^*$	$\max\{x_1, x_2\} - c_1 - c_2$ $\max\{x_1, x_2\} - c_1 - c_2$	$x_1 - c_1$ $x_1 - c_1 - c_2$

Table below shows payoffs of π_1 and π_2 contingent on realizations of x_1 and x_2.

Divided into four cells, depending on whether each x_i is below or above the cutoff x_2^*.

Within each cell

- upper rows are payoffs for strategy π_1
- lower rows are payoffs for strategy π_2
- we now cancel things that are the same for π_1 and π_2.

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	$-c_1$ 	$x_1 - c_1$ $x_2 - c_2$
$x_2 < x_2^*$		 $-c_2$

Now we replace c_1 and c_2.

$$x_2^* = -c_2 + \int \max\{x_2^*, x_2\}\, dF_2(x_2)$$
$$\implies -c_2 = \int \left\{ 1_{x_2 < x_2^*} 0 + 1_{x_2 \geq x_2^*} (x_2^* - x_2) \right\} dF_2(x_2).$$

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	$-c_1$	$x_1 - c_1$ $x_2 - c_2$
$x_2 < x_2^*$		$-c_2$

Now we replace c_1 and c_2.

$$x_2^* = -c_2 + \int \max\{x_2^*, x_2\}\, dF_2(x_2)$$
$$\implies -c_2 = \int \left\{ 1_{x_2 < x_2^*} 0 + 1_{x_2 \geq x_2^*} (x_2^* - x_2) \right\} dF_2(x_2).$$

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	$-c_1$	$x_1 - c_1$ x_2^*
$x_2 < x_2^*$		0

Now we replace c_1 and c_2.

$$
\begin{aligned}
x_1^* &= -c_1 + \int \max\{x_1^*, x_1\}\, dF_1(x_1) \\
\implies \quad -c_1 &= \int \Big\{ 1_{x_1 < x_1^*} 0 + 1_{x_1 \geq x_1^*} (x_1^* - x_1) \Big\}\, dF_1(x_1) \\
&\leq \int \Big\{ 1_{x_1 < x_2^*} 0 + 1_{x_1 \geq x_2^*} (x_2^* - x_1) \Big\}\, dF_1(x_1),
\end{aligned}
$$

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	$-\,c_1$	$x_1 - c_1$ x_2^*
$x_2 < x_2^*$		0

Now we replace c_1 and c_2.

$$
\begin{aligned}
x_1^* &= -c_1 + \int \max\{x_1^*, x_1\}\, dF_1(x_1) \\
\implies \quad -c_1 &= \int \left\{ 1_{x_1 < x_1^*} 0 + 1_{x_1 \geq x_1^*} (x_1^* - x_1) \right\} dF_1(x_1) \\
&\leq \int \left\{ 1_{x_1 < x_2^*} 0 + 1_{x_1 \geq x_2^*} (x_2^* - x_1) \right\} dF_1(x_1),
\end{aligned}
$$

	$x_1 < x_2^*$	$x_1 \geq x_2^*$
$x_2 \geq x_2^*$	0	x_2^*
	0	x_2^*
$x_2 < x_2^*$	0	0
	0	0

Done! π_2 is at least as good as π_1. $\square$

The proof in the general case is an easy induction with respect to the number of boxes which are still closed.

Fix a number of remaining boxes, and suppose that the Pandora rule is optimal when there are fewer than that number of remaining boxes.

Proof similar to that just completed shows that the first box opened should be the one with greatest reservation prize.

Of course we could have done the proof 'in algebra', but the figure makes it easy to follow.