Gittins Index, Pandora's Box, and Miller's Model of Learning and Labor Market Turnover

James J. Heckman
University of Chicago

Econ 350, Spring 2022

With material from:
Optimization Over Time - Dynamic Programming and Stochastic Control, Vol. 1
Peter Whittle, 1982

Gittins Index

Overview of the presentation

- General Model of Dynamic Discrete Choice
  - Solution: choose the project with highest Gittins index
- Three applications:

1. McCall's search model
2. Weitzman's Pandora's Box Model
3. Miller's Ocupational Choice Model

Objects of the Model

$C(X_t) = \{1, \ldots, N\}$: Choice set (projects)

$i_t \in C(X_t)$: Decision taken at time t

$X_t = \{X_t(1), \ldots, X_t(N)\}$: KN-dimensional structure. $X_t(i)$ is a K-dimensional vector describing project i at time t.

$r(X_t(i), i)$: Reward function for choosing i at time t. It depends only on $X_t(i)$ and not on states of other projects.

$p(X_{t+1}(i) \mid X_t(i), i)$: Markov transition probability. (Note: at any time, only state i changes. $X_{t+1}(j) = X_t(j)$, $j \neq i$. Only component of $X(t)$ changing is i at time t, $i = 1, \ldots, N$.)

Assume independence of projects

Value function:

$$V(X_t) = \sup_{i_j \in C_j(X)} \left\{ \sum_{j=t}^{\infty} \beta^{j-t} r(X_j(i), i_j) \right\}, \quad \forall\, j \geq t$$

Formal solution: Bellman's Equation

$$V(X_t) = \max_{1 \leq i \leq N} \left[r(X_t(i), i) + \beta \int V(X_t(1), \ldots, \overset{i^{\text{th}}}{\underset{\downarrow}{y}}, \ldots, X_t(N)) p(dy \mid X_t(i), i) \right]$$

Choice of i determines current reward and transition rate. Note that the i^{th} argument is the only one that changes if i is chosen.

Formulation

The problem defined in the last section is a Markov decision problem with the state variable $x = (x_1, x_2, \ldots, x_N)$. (Note that the subscript refers to the project, not to time). In order to keep within the classic framework of the discounted case we shall suppose rewards to be uniformly bounded:

$$-\infty < A(1-\beta) \leq R_i(x) \leq B(1-\beta) < \infty \tag{1}$$

(although see Exercise 4.1). The constants A and B are then lower and upper bounds on total discounted reward respectively.

For simplicity, let us denote the reward $R_{i(t)}(x_{i(t)}(t))$ realized at time t simply by $R(t)$. The total discounted reward is then $\sum_{t=0}^{\infty} \beta^t R(t)$ and its maximal expected value

$$G(x(0)) = \sup_{\pi} E_{\pi}\left[\sum_{t=0}^{\infty} \beta^t R(t) | W_0 \right] \tag{2}$$

is a function only of the initial state $x(0)$, as indicated. The reward function G will obey the dynamic programming equation

$$G = \max_{i} L_i G \tag{3}$$

and will be the unique bounded solution of this equation. Here L_i is the one-step operator if project i is engaged, so that

$$L_i G(x) = R_i(x_i) + \beta E[G(x(t+1)) | x(t) = x, i(t) = i]. \tag{4}$$

Of course, $x(t)$ and $x(t+1)$ differ only in the ith component.

Consider the modification of the process which allows the additional option of retirement for a lump reward of M. That is, one can either continue with selection from the projects or abandon operations permanently and settle for a retirement payment of M. The additional parameter M thus introduced turns out to be significant.

Let $\Phi(x, M)$ be the maximal expected conditional reward for the modified process, conditional on $x(0) = x$. Then Φ will be the unique bounded solution of the modified version of (3):

$$\Phi = \max\left[M, \max_i L_i \Phi \right]. \tag{5}$$

Let us refer to the original version of the process as the 'continuing process' and the modified version as the 'M process'. There is an immediate relation between the two.

Theorem 2.1. *$\Phi(x, M)$ is a non-decreasing convex function of M for which*

$$\Phi(x, M) = \begin{cases} G(x) & (M \leq A), \\ M & (M \geq B). \end{cases} \tag{6}$$

The optimal policies of the continuing process and of the M-process are identical for $M < A$.

Proof. To increase M cannot decrease reward, so the non-decreasing character is evident. To retire is always optimal for $M \geq B$ and to continue is always optimal for $M \leq A$ (for the M-process), whence (6) follows. To retire is never optimal if $M < A$, whence the last assertion.

Let V be the expected return from a policy whose prescription is independent of M. Then

$$V = V_c + ME(\beta^\tau) \tag{7}$$

where V_c is the expected reward before retirement (independent of M) and τ is the time of retirement. All expectations in (7) are those determined by the policy and are conditional on W_0. The event of non-retirement can be identified with the event $\tau = +\infty$, since, because $|\beta| < 1$, either convention will cause this contingency to yield a zero contribution to $E(\beta^\tau)$.

Since G is the supremum over policies of expression (7), linear in M, it is convex in M. ■

The Gittins Index

Suppose that to each project i can be attached an *index*, by which we mean a quantity $M_i(x_i)$ which is a function only of the label i and the current state x_i of that project. For simplicity we shall often denote the index for project i simply by M_i, but the dependence also on project state is understood.

A policy in which one always engages a project of currently greatest index will be referred to as an *index policy*. The conjecture that there might be an index policy which is optimal (among all policies) is an attractive one, since it asserts that optimal project choice reduces simply to a comparison of the figures-of-merit M_i. It is not obvious, however, that the conjecture is true.

In fact, Gittins (Gittins and Jones, 1974; Gittins, 1979) demonstrated that, with an appropriate choice of index, an index policy is optimal. The Gittins index is defined as follows. Let $\Phi_i(x_i, M)$ be the analogue of $\Phi(x, M)$ for the situation where, instead of the N projects, one has only the single project i. The sole options are then to continue with project i or to retire with reward M. The expected reward ϕ_i will then obey the reduced version of (2.5):

$$\phi_i = \max [M, L_i \phi_i]. \tag{1}$$

(Since L_i changes only the value of x_i its action on ϕ_i is defined.)

In analogy with the assertions of Theorem 2.1, $\phi_i(x_i, M)$ is a non-decreasing function of M which equals M for M large enough—certainly for $M \geq B$. The Gittins index $M_i(x_i)$ is defined as the infimal value of M for which $\phi_i(x_i, M) = M$. This is exactly the break-point at which one has $M = L_i\phi_i$. The index $M_i(x_i)$ can be regarded as the value of M which makes the options of continuation (with project i) or of retirement (with reward M) equally attractive.

$M_i(x_i)$ thus represents an equitable surrender price, or reserve price, for project i in its current state x_i. It is an equitable price under a particular circumstance, however. It is a price equivalent, not to the expected reward from indefinite operation of project i, but to the expected reward when one may continue with project i, but the option of accepting the price initially offered remains open at all times.

We shall refer to the index policy which uses the Gittins index as the *Gittins index policy*. Optimality of this policy will be proved in the next section. The index which Gittins actually used was

$$v_i = (1 - \beta)M_i$$

because he visualized the option of retiring at reward M as the option of adopting a *standard project* with a fixed reward rate $R = v$. The definitions differ only by a factor, and we shall find it convenient to work in terms of M rather than v.

Gittins has a second characterization of his index.

Theorem 3.1. *The Gittins index also has the characterization*

$$M_i(x_i) = \sup_{\pi} \frac{E_\pi[\sum_0^{\tau-1} \beta^t R_i(x_i(t)) | x_i(0) = x_i]}{1 - E_\pi(\beta^\tau | x_i(0) = x_i)} \tag{2}$$

or

$$v_i(x_i) = \sup_{\pi} \frac{E_\pi[\sum_0^{\tau-1} \beta^t R_i(x_i(t)) | x_i(0) = x_i]}{E_\pi[\sum_0^{\tau-1} \beta^t | x_i(0) = x_i]} \tag{3}$$

where the policy π determines a stopping time τ.

That is, v_i can be regarded as an average reward (with both reward and time discounted) over a stopping time τ, with the rule determining the stopping time chosen to maximize this average.

Proof. Since ϕ_i is the maximal reward under sequential choice between project i and retirement, then

$$\phi_i(x_i, M) \geq E_\pi\left[\sum_0^{\tau-1} \beta^t R_i(x_i(t)) + \beta^\tau M | x_i(0) = x_i\right]$$

for any π, with equality for some π, where τ is the instant of retirement. Giving M the value $M_i(x_i)$ we have then

$$M_i \geq E_\pi\left[\sum_0^{\tau-1} \beta^t R_i(x_i(t)) + \beta^\tau M_i | x_i(0) = x_i\right],$$

with equality for some π, whence (2) follows. ■

Optimality of the Gittins Index Policy

Define the function

$$\hat{\Phi}(x, M) = B - \int_M^B \prod_i \frac{\partial \phi_i(x_i, m)}{\partial m} \mathrm{d}m. \tag{1}$$

The motivation for this definition will transpire in the next section. Note that, since $\phi_i(x_i, m)$ is convex as a function of m, the derivative $\partial \phi_i / \partial m$ exists almost everywhere. The value assigned to it at the points of ambiguity will not affect the value of the integral (1), but for consistency we shall give it the value of the right-derivative.

It is now relatively easy to show that $\hat{\Phi} = \Phi$ and that this maximal reward is realized by the Gittins index policy, which is consequently optimal.

Lemma 4.1.

$$\hat{\Phi}(x, M) = \phi_i(x_i, M) P_i(x, M) + \int_M^\infty \phi_i(x_i, m) d_m P_i(x, m) \tag{2}$$

where

$$P_i(x, M) := \prod_{j \neq i} \frac{\partial \phi_j(x_j, M)}{\partial M} \tag{3}$$

is non-negative, non-decreasing in M and equal to unity for

$$M > M_{(i)} := \max_{j \neq i} M_i. \tag{4}$$

$$\hat{\Phi}(x, M) = \phi_i(x_i, M) P_i(x, M) + \int_M^\infty \phi_i(x_i, m) d_m P_i(x, m) \tag{2}$$

where

$$P_i(x, M) := \prod_{j \neq i} \frac{\partial \phi_j(x_j, M)}{\partial M} \tag{3}$$

is non-negative, non-decreasing in M *and equal to unity for*

$$M > M_{(i)} := \max_{j \neq i} M_i. \tag{4}$$

Proof. Equation (2) follows from (1) by partial integration. Since ϕ_i, as a function of M, is non-decreasing, convex and equal to M for $M \geq M_i$, then $\partial \phi_i / \partial M$ is non-negative, non-decreasing and equal to unity for $M \geq M_i$. The properties asserted for P_i thus follow. ■

Consider the quantity

$$\delta_i(x_i, M) = \phi_i(x_i, M) - L_i\phi_i(x_i, M)$$

and note that $\delta_i \geq 0$, with equality for $M \leq M_i$. We are interested for the moment

in the dependence of the various quantities on M for fixed x, and so shall for simplicity suppress the x-argument.

$$\hat{\Phi} \geq M, \tag{5}$$

with equality if $M \geq \max M_j$, *and*

$$\hat{\Phi}(M) - L_i \hat{\Phi}(M) = \delta_i(M) P_i(M) + \int_M^\infty \delta_i(m) d_m P_i(m) \geq 0, \tag{6}$$

with equality if $M_i = \max M_j \geq M$.

Proof. Inequality (5) and the characterization of the equality case follow from (2) and the properties of P_i.

The first relation of (6) follows immediately from (2). The non-negativity of the expression follows from $\delta_i(M) \geq 0$ and the non-negative and non-decreasing nature of P_i. We know that $\delta_i(M) = 0$ for $M \leq M_i$ and that $d_m P_i(m) = 0$ for $m \geq M_{(i)}$, so that expression (6) will be zero if $M \leq M_i$ and $M_{(i)} \leq M_i$. This pair of conditions is equivalent to those asserted in the Lemma. ■

Theorem 4.1. $\hat{\Phi} = \Phi$ *and the Gittins index policy is optimal.*

Proof. We shall have demonstrated that $\hat{\Phi} = \Phi$ if we can show that $\hat{\Phi}$ satisfies (2.5), because Φ is the unique bounded solution of this equation (Theorem 23.3.1). We shall have demonstrated that the Gittins index policy is optimal if we can show that the maximizing option in the right-hand member of (2.5) is just that option which would be recommended by the Gittins index policy (Theorem 23.3.2). But the truth of both of these statements follows from the inequalities and cases of equality proved in Lemma 4.2. ■

Our proof that $\hat{\Phi} = \Phi$ establishes the truth of the identity

$$\Phi(x, M) = B - \int_M^B \prod_i \frac{\partial \phi_i(x_i, m)}{\partial m} \mathrm{d}m, \tag{1}$$

and the definition (4.1) of $\hat{\Phi}$ has its origin in the conjecture that identity (1) holds. We now outline the motivation for this conjecture.

Lemma 5.1. *Let V be the expected reward, conditional on W_0, for an arbitrary policy whose prescription is independent of M. Then*

$$\frac{\partial V}{\partial M} = E(\beta^{\tau} | W_0) \tag{2}$$

where τ is the time of retirement under this policy.

Proof. This is an immediate consequence of (2.7). ■

Let us now define a *write-off policy* as a policy in which project i is written off (i.e. abandoned) when first its state x_i enters a *write-off set* S_i $(i = 1, 2, \ldots, N)$. One continues as long as there are projects which have not been written off, working only on those projects; one retires as soon as all projects are written off. That is, one continues until all projects have been driven into their write-off sets.

Note that a write-off policy is not fully specified by prescription of the write-off sets S_i, because no rule has been given for the order in which active projects are to be engaged. This is something the write-off policy does not specify; moreover, a write-off policy need be neither Markov nor stationary. Note that the Gittins index policy is a write-off policy, with S_i the set of x_i for which $M_i(x_i) \leq M$.

Lemma 5.2. *Let τ be the time to retirement for an N-project policy, and τ_i the time to retirement when only project i is available. If the policy is a write-off policy, then*

$$E[\beta^{\tau} | W_0] = \prod_i E[\beta^{\tau_i} | W_0]. \tag{3}$$

Proof. We have $\tau = \sum \tau_i'$ where τ_i' is the total *process time* for project i, the total time for which project i is engaged before it is written off. Equation (3) states effectively that the τ_i' are independently distributed and that τ_i, τ_i' have the same distribution (conditional on W_0). But since evolution of the projects (in process time) is independent, the process time needed to take x_i into S_i is independent of the states of other projects; hence the assertion. ■

Lemma 5.3. *Let V and V_i be the expected rewards, conditional on W_0, for an arbitrary write-off policy whose prescription is independent of M when N projects are available and when only project i is available respectively. Then*

$$\frac{\partial V}{\partial M} = \prod_i \frac{\partial V_i}{\partial M}. \tag{4}$$

Proof. Relation (4) follows from (2), (3). Note Exercise 1. ■

Suppose now that relation (4) also holds for the optimal policies, so that

$$\frac{\partial \Phi}{\partial M} = \prod_i \frac{\partial \phi_i}{\partial M}. \tag{5}$$

Identity (5) would then follow immediately by integration of (5) and appeal to $\Phi(x, B) = B$. To establish (5) we would need analogues of Lemmas 5.1 and 5.2 for optimal policies. Lemma 5.1 indeed has an analogue:

Lemma 5.4. *Let E_M denote expectation under a policy optimal when retirement reward has the value M. Then $E_M(\beta^\tau | x(0) = x)$ is a sub-gradient of $\Phi(x, M)$, as a function of M, and*

$$\frac{\partial \Phi(x, M)}{\partial M} = E_M[\beta^\tau | x(0) = x] \tag{6}$$

Proof. Since in each state there are at most $N+1$ possible actions, optimal policies always exist. If one applies an M-optimal policy to the $(M+\delta)$-process one achieves an expected reward not exceeding $\Phi(x, M+\delta)$. This assertion and

relation (2.7) imply that, for any M, δ,

$$\Phi(x, M+\delta) - \Phi(x, M) \geq \delta E_M(\beta^\tau | x(0) = x). \tag{7}$$

(The expectation is conditional on W_0, but $x(0)$ will be the only effective conditioner for an optimal policy.) Relation (7) characterizes expression (6) as a sub-gradient to Φ. It will then coincide with the gradient of Φ, wherever this exists. But Φ, being convex, will have a gradient for almost all M, whence the assertion of the Lemma. ■

However, there will be no analogue of Lemma 5.2 unless one can assert that there is an optimal policy (for fixed M) which is a write-off policy. This is the real conjecture, which is plausible, but for which there seems to be no direct proof. Its value is that it leads to the conjecture $\hat{\Phi} = \Phi$, whose truth we established in Theorem 4.1. This theorem also established optimality of the Gittins index policy, so that there is indeed a write-off policy which is optimal.

Example: The Deteriorating Case

To complete the treatment for individual cases one must now calculate the index $M_i(x_i)$ from one of the two characterizations of Section 3. Relatively few cases are known in which an explicit formula for the index can be deduced, so one will often have to resort to numerical solution of the optimality equation (3.1) for an appropriate range of values of M. However, even if one is driven to numerical solution at this point, the index result still represents an enormous reduction of the original problem, as well as an essential insight.

In this and the following sections we review a number of cases for which the index can be determined explicitly. The material of Sections 6 to 9 is due to Gittins (1979).

Theorem 6.1 (The deteriorating case). *Suppose that* $\phi_i(x_i(t+1), M) \leq \phi_i(x_i(t), M)$ *for all* M *and for all* $(x_i(t), x_i(t+1))$ *which occur with positive probability. Then*

$$M_i(x_i) = \frac{R_i(x_i)}{1-\beta}. \tag{1}$$

If this holds for all i then the Gittins rule is a one-step look-ahead rule.

Interpretations and Extensions

Define

$$\nu_i(X_i) = (1-\beta)M_i$$

In the case of no uncertainty

$$\nu_i(X_i) = \sup_{\tau} \frac{\left(\sum_{t=0}^{\tau-1} \beta^t R_i(X_i(t))\right)}{\sum_{t=0}^{\tau-1} \beta^t}.$$

Gittins Index is a time-discounted average return.

An alternate to Gittins Index is

$$\nu_i(X_i) = \sup \frac{E\left(\sum_{j=0}^{\tau-1} \beta^j R(X_j(i), i) \mid X_0(i) = X_i\right)}{E\left(\sum_{j=0}^{\tau-1} \beta^j \mid X_0(i) = X_i\right)}.$$

Applications: Job Search, Pandora's Box and Miller**Job Search:**

(One Spell Model)

- Jobs last forever, no learning.
- Only issue is when to stop.
- This is a degenerate, single arm bandit.
- $c =$ cost of playing machine.
- $X_t =$ reward on t^{th} trial.
- No learning implies X_t is independent and identically distributed with cdF $F(x)$ known to agent.

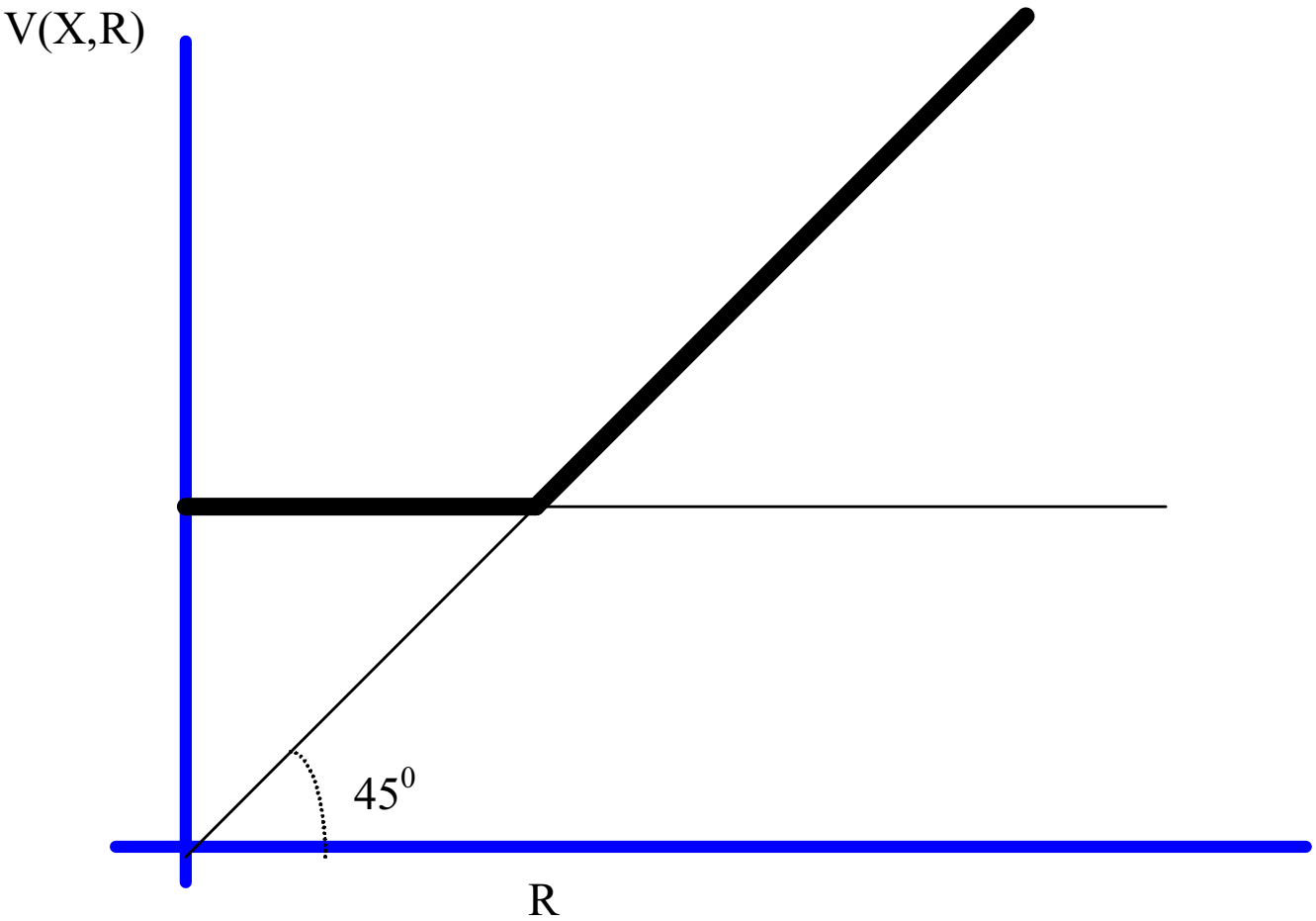
$$V(X) = \max\left[X, -c + \beta \int_0^\infty V(y)dF(y)\right] \tag{1}$$

Optimal policy: Search if $X \in [0, R]$ reservation price is R :

$$V(X) = \begin{cases} R = -c + \beta \int_0^\infty V(y)dF(y) & \text{if } X < R \\ X & \text{if } X \geq R \end{cases} \tag{2}$$

R is the value that makes a person indifferent between stopping and continuing.

This figure graphs the functional equation (1) and it reveals that the optimal solution is of the form of (2) (see e.g. Sargent's textbook Ch. 6):



Solving for reservation wage:

using (2) we convert (1) into an ordinary equation in the reservation wage:

$$R = -c + \beta \left[\int_0^R R dF(x) + \int_R^\infty X \, dF(x) \right]$$

$$R \left(\int_0^R dF(x) + \int_R^\infty dF(x) \right) = -c + \beta \left[\int_0^R R dF(x) + \int_R^\infty X \, dF(x) \right]$$

or

$$(1-\beta) R \int_0^R dF(x) = -c + \int_R^\infty (\beta X - R) \, dF(x)$$

Solving for reservation wage:

adding $(1-\beta)R\int_R^\infty dF(x)$ to both sides we have

$$R = \frac{-c}{1-\beta} + \frac{\beta}{1-\beta}\int_R^\infty (X-R)\,dF(x) \tag{3}$$

with unique solution for R. To see this, define

$$g(R) = \frac{-c}{1-\beta} + \frac{\beta}{1-\beta}\int_R^\infty (X-R)\,dF(x)$$

Solving for reservation wage:

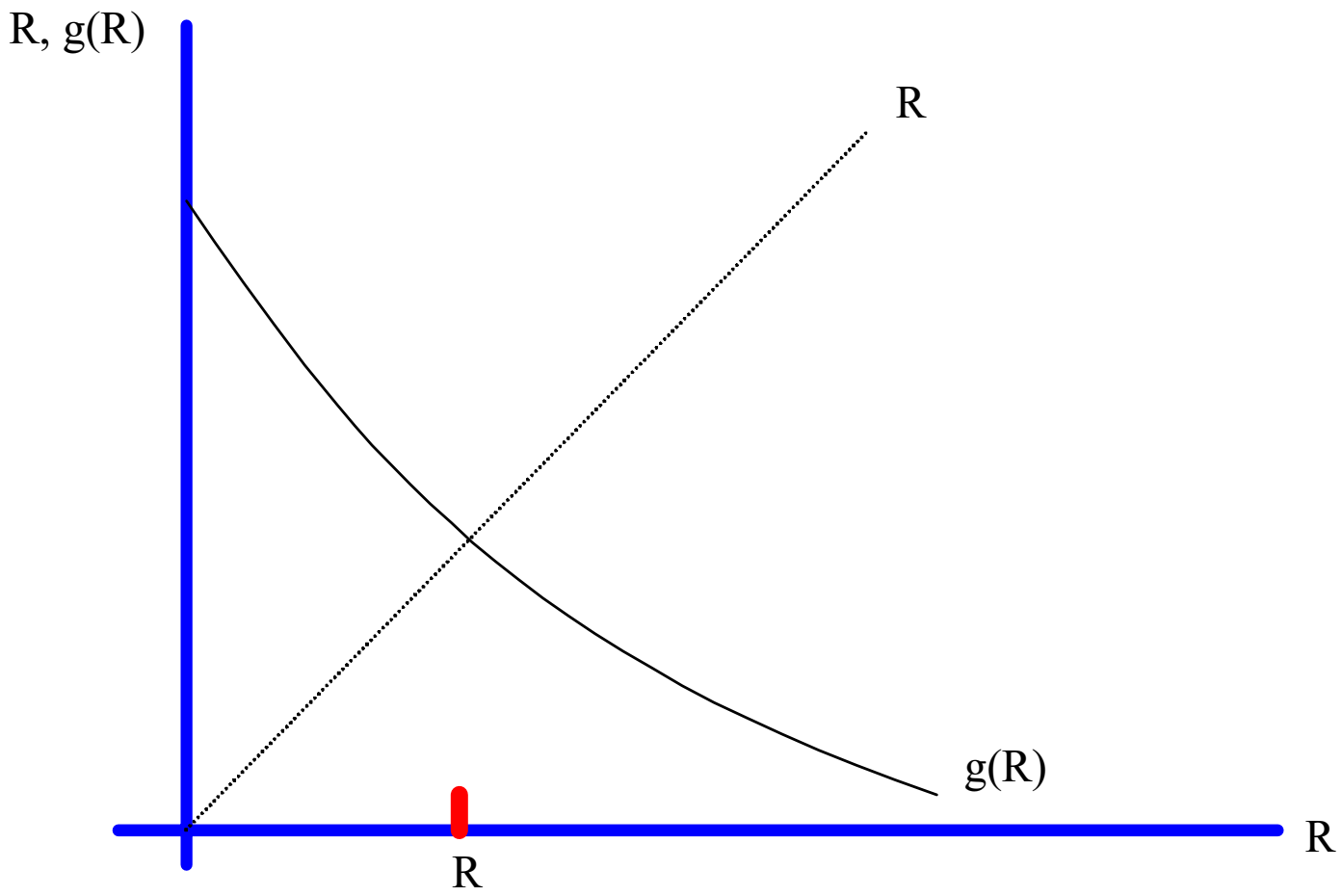
with

$$
\begin{aligned}
g'(R) &= -\frac{\beta}{1-\beta}\left[1-F(R)\right] < 0 \\
g''(R) &= \frac{\beta}{1-\beta}F'(R) > 0
\end{aligned}
$$

the optimal reservation wage satisfies, from (3)

$$R^* = g(R^*)$$

Solving for reservation wage:



Pandora (Weitzmann, *Econometrica*, May, 1979)

- N different occupations (or N college majors) each yielding an unknown reward
- Each occupation has its own search cost c_i and independent probability distribution F_i for the reward X_i
- Occupations are sampled sequentially, in whatever order is desired. When it has been decided to stop searching, only one occupation is accepted, the maximum sampled reward.
- Under this formulation, what sequential search strategy maximizes expected present discounted value?

- **Pandora's problem:** At each state Pandora must decide wheter or not to open a box. If she chooses to stop searching, Pandora collects at that time the maximum reward she has thus far uncovered. Should Pandora wish to continue sampling, she must select the next box to be opened and pay c_i
- **Solution:**

- Compute Reservation Price for each box
- Try jobs with highest R_i^*'s
- Keep trying until one gets $X_t(i^*) \geq \max_i\{R_i^*\}$

(*i.e.* keep going until you get a realization $\geq$ sampled values).

Proof: Compute reservation prices for each occupation $R_i(x_i)$. Consider occupation i. If occupation i tried, then X_i is known. For an i-th occupation with known trial,

$$V_i(X_i, R) = \max[R, X_i].$$

Therefore, $R_i(X_i) = X_i$. Thus, the reservation price for a sampled occupation is just the wage realized.

For an untried occupation,

$$\begin{aligned}
\widetilde{V}_i(R) &= V_i(-c_i, R) \\
&= \max\left[R, -c_i + \beta \int_0^\infty V_i(y, R)\, dF_i(y)\right] \\
&= \max\left[R, -c_i + \beta \left(F_i(R)R + \int_R^\infty y\, dF_i(y)\right)\right]
\end{aligned}$$

Reservation wage is the smallest value of R_i such that

$$R_i = -c_i + \beta R_i F_i(R_i) + \beta \int_R^\infty y \, dF_i(y)$$

$$\therefore R_i = -\frac{c_i}{1-\beta} + \frac{\beta}{1-\beta} \int_{R_i}^\infty (X - R_i) \, dF_i(y)$$

as in the above search model.

Consider a binomial case with $\beta = 1$ (no discounting):

$$X_i = 0 \quad \text{wp.} \quad (1 - p_i)$$

$$X_i = r_i \quad \text{wp.} \quad p_i$$

and R_i satisfies

$$c_i = \int_{R_i}^{\infty} (X - R_i)\, dF_i(X)$$

2 cases:

- If $R_i = 0$, then

$$c_i = \int_0^\infty X \, dF_i(X) = p_i r_i.$$

No reason for this to be satisfied in general. Therefore,

- We assume $R_i \in (0, r_i]$

$$c_i = \int_{R_i}^\infty (X - R_i) \, dF_i(X) = p_i r_i - R_i p_i$$

$$\therefore \quad R_i = \frac{p_i r_i - c_i}{p_i}$$

- Thus, take two projects with equal expected value ($p_i r_i - c_i$). The project with lower expected probability of reward is the one to try (go for riskier project). Why? Suppose costs are the same. Then r_i is higher, since p_i is lower.
- Suppose that rewards are the same. But costs being lower implies that p_i is lower (trying less costly combo).
- No further learning.

Theorem: Suppose that $X_1, \ldots, X_n$ is a r.s. from a normal distribution with unknown mean W and precision $r = 1/\sigma^2$. Suppose that the prior distribution of W is a normal with mean μ and precision $\tau = 1/\sigma_p^2$, $-\infty < W < \infty$, $0 < \sigma^2 < \infty$, and $0 < \sigma_p^2 < \infty$.Then the posterior of W when $X = X_i$ $(i = 1, \ldots, n)$ is normal with mean μ' and precision $\tau + nr$, where

$$\mu' = \frac{\tau\mu + nr\overline{X}}{\tau + nr}.$$

Proof: (obvious)
The likelihood of the normal r.s. $X_1, ..X_n$ is proportional to the likelihood of the sample mean $\overline{X}$

$$\begin{aligned} f_n(X_1, \ldots, X_n/W) &\propto \exp\left[-\frac{r}{2}\sum_{i=1}^{n}(X_i - W)^2\right] \\ &\propto \exp\left[-\frac{nr}{2}(\overline{X} - W)^2\right] \end{aligned}$$

Prior

$$f(W) \propto \left[-\frac{\tau}{2}(W - \mu)^2\right]$$

Posterior proportional to prior times likelihood:

$$\begin{aligned}
\tau(W-\mu)^2 + nr(W-\overline{X})^2 &= \text{completing the squares....} \\
&= (\tau+nr)(W-\mu')^2 + \frac{\tau nr(\overline{X}-\mu)^2}{\tau+nr} \\
f(W \mid \overline{X}) &\propto \exp\left[-\left(\frac{\tau+nr}{2}\right)(W-\mu')^2\right]
\end{aligned}$$

$$\mu' = \frac{\tau\mu + nr\overline{X}}{\tau+nr}$$

As $n \to \infty$, we weight sample mean more and more strongly over time.

Note

$$
\begin{aligned}
E(\mu') &= \frac{\tau\mu + nrW}{\tau + nr} \\
\mathrm{Var}(\mu') &= \left(\frac{\tau}{\tau + nr}\right)^2 \overbrace{\mathrm{Var}(\mu)}^{=0} + \left(\frac{nr}{\tau + nr}\right)^2 \mathrm{Var}(\overline{X}) \\
&= \left(\frac{nr}{\tau + nr}\right)^2 \left(\frac{1}{nr}\right) \\
&= \frac{nr}{(\tau + nr)^2}
\end{aligned}
$$

As $n \to \infty$, $\mathrm{Var}(\mu') \to 0$. Monotonically decreasing in n.

Miller's Model

- N independent occupations
- Reward : $r_t(i) = \eta_i + \sigma(i)\varepsilon_t(i)$. i=1,...,N
- When we try each occupation we learn more. We assume $\varepsilon_t(i)$ is independent and identically distributed normal.

Suppose that worker's prior beliefs about match η_i are normally distributed.

$$\eta_i \sim N(\gamma_0(i), \beta_0^2(i))$$

Precision is

$$\overline{P}_0(i) = \frac{1}{\beta_0^2(i)}.$$

After one trial on the job, the posterior mean is

$$\begin{aligned} \gamma_1(i) &= \frac{\gamma_0(i)\bar{P}_0(i) + P_{\sigma^2} r_1(i)}{\overline{P}_0(i) + P_{\sigma^2}} \\ \overline{P}_{(1)}(i) &= \overline{P}_0(i) + \overline{P}_{\sigma^2}. \end{aligned}$$

Keep going. After τ periods on the job,

$$\gamma_\tau(i) = \frac{\gamma_0(i)\bar{P}_0(i) + (\bar{r}_\tau)(\tau P_{\sigma^2})}{\bar{P}_0(i) + (\tau P_{\sigma^2})}$$

where

$$\begin{aligned} \bar{r}_\tau(i) &= \frac{1}{\tau}\sum_{j=1}^{\tau} r_j(i) \qquad \tau > 0 \\ \bar{P}_{(\tau)}(i) &= \bar{P}_0(i) + \tau P_\sigma \end{aligned}$$

Note several features. How much learning takes place?

$$\begin{aligned}
\gamma_\tau &= \frac{\gamma_0(i)\overline{P}_0(i) + \left(\sum_{j=1}^{\tau-1} r_j(i)\right) P_{\sigma_0^2}}{\overline{P}_0(i) + \tau P_{\sigma_0^2}} + \frac{r_\tau(i) P_{\sigma_0^2}}{\overline{P}_0(i) + \tau P_{\sigma^2}} \\
&= \left[\frac{\overline{P}(i) + (\tau-1) P_{\sigma_0^2}}{\overline{P}_0(i) + \tau P_{\sigma_0^2}}\right] \gamma_{\tau-1}(i) + \frac{r_\tau(i) P_{\sigma_0^2}}{\overline{P}_0(i) + \tau P_{\sigma^2}} \\
&= \left(1 - \frac{P_{\sigma_0^2}}{\overline{P}_0(i) + \tau P_{\sigma_0^2}}\right) \gamma_{\tau-1}(i) + \left(\frac{P_{\sigma_0^2}}{\overline{P}_0(i) + \tau P_{\sigma^2}}\right) r_\tau(i)
\end{aligned}$$

Change in mean small. Thus, learning vanishes (noise component being small).

- Now the model is in the form of a Gittins Index Problem. Theorem 1 applies. Every period choose the job with highest R_i
- We update density $P(dy \mid X(t))$.

$$X(t) = [\gamma_\tau(i), \bar{P}_\tau(i)].$$

- Therefore, Markov full state description.
- Conditional distribution of $\gamma_\tau \mid \gamma_{\tau-1}$ is

$$N\left(\gamma_\tau, \left(\frac{\tau P_{\sigma^2}}{[\bar{P}_0(i) + \tau P_{\sigma^2}]^2}\right)\right)$$

Then, the value function at time t for a given R is

$$V(X(t), R) = \max[R, \int_{-\infty}^{\infty} V(X(t+1), R) P(\cdot \,/\, X(t))]$$

where

$$\begin{aligned} X(t) &= [\gamma_t(i), \bar{P}_t(i)] \\ V(X(t), R) &= \max \left[\begin{array}{c} R, \int_{-\infty}^{\infty} V\left(\left(1\text{-} \left[\frac{P_{\sigma_0^2}}{\bar{P}_0(i) + t P_{\sigma_0^2}} \right] \right) \gamma_t(i) + \\ \left(\frac{P_{\sigma_0^2} r_{t+1}(i)}{\bar{P}_0(i) + t P_{\sigma_0^2}} \right), \bar{P}_0(i) + (t+1) P_{\sigma_0^2}, R \right) f(r)\, dr \end{array} \right] \end{aligned}$$

where r is normally distributed:

Procedure

For a fixed R, solve for V (solve fnl. equation). Then find

$$R = \beta \int_{-\infty}^{\infty} V(X(t+1, R) P(\cdot \mid X(t))$$

- Intuitively, the Gittins index for any job R_i is just the sum of its expected return plus a term which represents the value of the job as a source of information about itself.
- The algorithm for computing R_i is based in 2 stages: first find the value function and then find R_i

- Take a simple search model.
- What is the expected value of the search?
- The only cost is the time foregone:

$$V = E\max\left[\frac{x}{r}, V\right]$$

$$= \beta\left[V\int_{-\infty}^{V} dF(x) + \int_{V}^{\infty} x\, dF(x)\right]$$

$$= \beta V + \beta\int_{V}^{\infty} (x - V)\, dF(x)$$

$$V(1-\beta) = \beta \int_V^\infty (x - V)\, dF(x)$$

- When you raise the value of the search, you raise

$$rV = \frac{\beta}{1-\beta} \int_V^\infty (x - V)\, dF(x).$$

- Take a normal distribution and add variance to it.
- You raise the value of the search, V.
- If you reduce the variance, you reduce the value of the search.
- Note that as you pull mass toward the upper tail, you are better off!

- But as you learn more, reserve value shrinks:

$$rV_t = \left(\frac{\beta}{1-\beta}\right) \int_{rV_t}^{\infty} (X - rV_t)\ dF_t(x).$$

- Thus with learning, $V_t \downarrow$, but the mean doesn't change much at all.
- You shop around less and less.

- Take a population with mean and variance give by 1.

- $$Y_n = N\left(0, \frac{\tau + n}{\tau + n - 1}\right)$$

$$\sigma^2 = \left(\frac{\tau + n}{\tau + n - 1}\right)$$

$$\frac{\partial \sigma^2}{\partial \tau} = \frac{1}{\tau + n - 1} - \frac{\tau + n}{(\tau + n - 1)^2}$$

$$= \frac{(\tau + n - 1) - (\tau + n)}{(\tau + n - 1)^2}$$

$$= \frac{-1}{\tau + n - 1} < 0$$

- Thus,

$$\tau \uparrow \Rightarrow \sigma^2 \downarrow .$$

- Raise n to get

$$\frac{\partial \sigma^2}{\partial n} = \frac{1}{\tau + n - 1} - \frac{\tau + n}{(\tau + n - 1)^2} = \frac{-1}{(\quad)^2} < 0.$$

- Thus variance of Y_n is decreasing in τ and n:

$$\sigma^2_{Y_N} \downarrow \quad \text{as} \quad n \uparrow .$$