Optimal Stopping of Markov chain, Gittins Index and Related Optimization Problems

Isaac M. Sonin

Department of Mathematics and Statistics
University of North Carolina at Charlotte, USA

http://...Isaac Sonin in Google

New York, Columbia University, September 2011
Outline

- Optimal Stopping (OS) of Markov Chains
- State Elimination (SE) Algorithm
- Gittins, Katehakis - Veinott and w (Whittle) Indices and their Generalizations
- Three Abstract Optimization Problems
  - Abstract Optimization Equality
- Continue, Quit, Restart Probability Model
- Open Problems
There are two approaches - "Martingale theory of OS "and "Markovian approach".

Classical monographs:

- Chow, Robbins and Sigmund (1971)
- Dynkin, Yushkevich (1967)
- T. Ferguson (website)
T. Ferguson: "Most problems of optimal stopping without some form of Markovian structure are essentially untractable...”.

OS Model $M = (X, P, c, g, \beta)$: continue or stop

- $X$ finite (countable) state space,
- $P = \{p(x, y)\}$, stochastic (transition) matrix
- $c(x)$ one step cost function,
- $g(x)$ terminal reward function,
- $\beta$ discount factor, $0 \leq \beta \leq 1$
- $(Z_n)$ MC from a family of MCs defined by a Markov Model $M = (X, P)$
- $v(x) = \sup_{\tau \geq 0} E_x \left[ \sum_{i=0}^{\tau-1} \beta^i c(Z_i) + \beta^\tau g(Z_\tau) \right]$ value function
Remark! absorbing state \( e, \ p(e, e) = 1 \),
\[
p(x, y) \rightarrow \beta p(x, y), \ p(x, e) = 1 - \beta,
\]
\[
\beta \rightarrow \beta(x) = P_x(Z_1 \neq e) \text{ probability of } "\text{survival}."
\]

- \( S = \{ x : g(x) = v(x) \} \) optimal stopping set.
- \( Pf = Pf(x) = \sum_y p(x, y) f(y) \).

Theorem (1, Shiryayev 1969)

(a) The value function \( v(x) \) is the minimal solution of Bellman equation ...

\[
v = \max(g, c + Pv),
\]

(b) if state space \( X \) is finite then set \( S \) is not empty and
\[
\tau_0 = \min\{ n \geq 0 : Z_n \in S \} \text{ is an optimal stopping time. } ...
\]
Basic methods of solving OS of MC, \( c \equiv 0 \)

- The direct solution of the Bellman equation
- The value iteration method: one considers the sequence of functions
  \[ v_n(x) = \sup_{0 \leq \tau \leq n} E_x \ldots, v_{n+1}(x) = \max(g(x), Pv_n(x)), v_0(x) = g(x). \]
  Then \( v_0(x) \leq v_1(x) \leq \ldots v_n(x) \) converges to \( v(x) \).
- The linear programming approach (\(|X| < \infty\)),
  \[ \min \sum_{y \in X} v(y), \quad v(x) \geq \sum_y p(x, y)v(y), \quad v(x) \geq g(x), \quad x \in X. \]
- Duality Theory, Harmonic function method (Haugh & Kogan, S. Christensen)
State Elimination Algorithm for OS of MC

OS = Bellman equation \( v(x) = \max(g(x), c(x) + P v(x)) \);

\( M_1 = (X_1, P = P_1, c = c_1, g), S = S_1 \). Three simple facts:

1. It may be difficult to find the states where it is optimal to stop, \( g(x) \geq c_1(x) + P_1 v(x) \), but it is easy to find a state (states) where it is optimal not to stop: do not stop if \( g(z) < c_1(z) + P_1 g(z) \leq c_1(z) + P_1 v(z) \).

2. After identifying these states, set \( G \), we can "eliminate" the subset \( D \subset G \), and recalculate \( P_1 \rightarrow P_2 \) and \( c_1 \rightarrow c_2, g \). Elimination theorem: \( S_1 = S_2, v_1 = v_2 \). Repeat these steps until \( g(x) \geq c_k(x) + P_k g(x) \) for all remaining \( x \in X_k \). Then

3. Proposition 1. Let \( M = (X, P, c, g) \) be an optimal stopping problem, and \( g(x) \geq c(x) + Pg(x) \) for all \( x \in X \). Then \( X \) is the optimal stopping set in the problem \( M \), and \( v(x) = g(x) \) for all \( x \in X \).
Eliminate state(s) \( z \), (set \( D \)) and recalculate probabilities

*Embedded Markov chain* (Kolmogorov, Doeblin) \( M_1 = (X_1, P_1) \), \( D \subset X_1 \), \( X_2 = X_1 \setminus D \), \((Z_n)\) MC \( \tau_0, \tau_1, \ldots, \tau_n, \ldots \), the moments of zero, first, and so on, visits of \((Z_n)\) to the set \( X_2 \). Let \( Y_n = Z_{\tau_n}, n = 0, 1, 2, \ldots \)

**Lemma (KD)**

(a) The random sequence \((Y_n)\) is a Markov chain in a model \( M_2 = (X_2, P_2) \), where \( P_2 = \{p_2(x, y)\} \) is given by formula

\[
P_1 = \begin{bmatrix} Q & T \\ R & P_0 \end{bmatrix}, \quad P_2 = P_0 + RU = P_0 + RNT,
\]

\( N \) is a (transient) fundamental matrix, i.e. \( N = (I - Q)^{-1} \),

\[
N = I + Q + Q^2 + \ldots = (I - Q)^{-1}, \quad U = NT.
\]

State Reduction Approach: GTH/S algorithm to calculate the invariant distribution (1985) for \( D = \{z\} \)
State Elimination Algorithm, \( c \equiv 0 \)

If \( D = \{z\} \) then

\[
p_2(x, y) = p_1(x, y) + p_1(x, z)n_1(z)p_1(z, y),
\]

where \( n_1(z) = 1/(1 - p_1(z, z)) \).

State Elimination Algorithm (for \( c(x) = 0 \) )

\[
g(x) - (P g(x) + c(x)) = g - P g
\]

\[
g(x) - P_1 g(x) \geq 0 \quad \text{for all } x
\]

\[
X_1 = S
\]

there is \( z : g(z) - P_1 g(z) < 0 \)

\[
M_1 \rightarrow M_2 : g(x) - P_2 g(x)
\]

... and so on
Some References

**State Reduction approach**
GTH/S algorithm (1985), invariant distr. for ergodic MC;
W. Grassmann, M. Taksar, D. Heyman, (1985),

**State Elimination for OS of MCs**
Presman, E., Sonin, I. (1972). The problem of best choice ... a random number of objects.


E. Presman (2011) Continuous time OS
There is a well known connection between three problems related to Optimal Stopping of Markov Chain and the equality of three corresponding indices: the classical Gittins index in the Ratio Maximization Problem, the Kathehakis-Veinot index in a Restart Problem, and \( w \) (Whittle) index in a family of Retirement Problems.

from these indices \( \rightarrow \) to generalized indices \( \rightarrow \) to ...

One of the goals of my talk is to demonstrate that the equality of these (generalized) indices is a special case of a more general relation between three simple abstract optimization problems.

There is no doubt that the relationship between these problems was used in optimization theory before on different occasions in specific problems but we fail to find a general statement of this kind in the vast literature on optimization.
Three indices for MC reward model. Gittins index

**Reward Model** $M = (X, P, c(x), \beta)$, *continue or stop.*

Given a reward model $M$ and point $x \in X$, the *classical Gittins index*, $\gamma(x)$, is defined as the *maximum of the expected discounted total reward* during the interval $[0, \tau)$ *per unit of expected discounted time* for the Markov chain starting from $x$, i.e.

$$
\gamma(x) = \sup_{\tau > 0} \frac{E_x \sum_{n=0}^{\tau-1} \beta^n c(Z_n)}{E_x \sum_{n=0}^{\tau-1} \beta^n}, \quad 0 < \beta = \text{const} \leq 1.
$$

*Multi-armed bandit (MAB) Problems:* a number of competing projects, each returning a stochastic reward. Projects are *independent* from each other and only one project at time may evolve.

*Gittins Theorem:* Gittins index policy is optimal.


KV index

KV index $M = (X, s, P, c(x), \beta)$, continue or restart to $s$.

Let $h(x|s)$ denote the supremum over all strategies of the expected total discounted reward on the infinite time interval in reward model with an initial point $x$, and restart point $s$. Using the standard results of Markov Decision Processes theory, Katehakis and Veinott proved that function $h(x|s)$ satisfies the equality

$$h(x|s) = \sup_{\tau > 0} E_x\left[ \sum_{n=0}^{\tau-1} \beta^n c(Z_n) + \beta^\tau h(s) \right],$$

and $\gamma(s) = (1 - \beta)h(s)$, where by definition $h(s) = h(s|s)$. We call index $h(s)$ a KV index. This index can be defined for any point $x \in X$, so we use also notation $h(x)$. 

Isaac M. Sonin
Optimal Stopping of Markov chain, Gittins Index and Related Optimization Problems
New York, Columbia University, September 2011
**w (Whittle) index**

*Retirement Process* formulation was provided by Whittle (1980). Given a reward model $M$, he introduced the parametric *family of OS models* $M(k) = (X, P, c(x), k, \beta)$, where parameter $k$ is a real number and the terminal reward function $g(x) = k$ for all $x \in X$.

Denote $v(x, k)$ the value function for such a model, i.e.

$$v(x, k) = \sup_{\tau \geq 0} E_x[\sum_{n=0}^{\tau-1} \beta^n c(Z_n) + \beta^\tau k],$$

and denote Whittle index

$$w(x) = \inf\{k : v(x, k) = k\}.$$ 

Since $\beta < 1$, for sufficiently large $k$ it is optimal to stop immediately and $v(x, k) = k$. Thus $w(x) < \infty$. The results of Whittle imply that $v(x, k) = k$ for $k \geq w(x)$, $v(x, k) > k$ for $k < w(x)$, and $w(x) = h(x)$. 
Theorem 2

The three indices defined for a reward model
\[ M = (X, P, c(x), \beta), 0 < \beta < 1, \text{ coincide, i.e.} \]
\[ h(x) = w(x) = \gamma(x)/(1 - \beta), x \in X. \]

Sonin (Stat. & Prob. Let., 2008): simple and transparent algorithm to calculate this common index. This algorithm is based on State Elimination algorithm. Nino-Mora (2007) for classical GI.

To apply this algorithm it is necessary to replace a constant discount factor \( \beta \) by a variable ”survival” probability \( \beta(x) \), because after the first recursive step a discount factor is not a constant anymore. So by necessity a more general model was considered and the classical GI \( \gamma(x) \) was replaced by a generalized Gittins Index (GGI) \( \alpha(x) \) as follows.
Generalized Gittins index

In general case, when $\beta(x)$ can be variable, we denote $P_x(Z_\tau = e)$ by $Q^\tau(x)$, the probability of termination on $[0, \tau)$, and we define the Generalized GI (GGI), $\alpha(x)$, for a model with termination as

$$\alpha(x) = \sup_{\tau > 0} \frac{R^\tau(x)}{Q^\tau(x)},$$

i.e. $\alpha(x)$ is the maximum discounted total reward per chance of termination.

Mitten, L. G. (1960); Denardo, Eric V.; Rothblum, Uriel G.; Van der Heyden, Ludo (2004) Index policies for stochastic search in a forest ... Math. Oper. Res.

The common part of all three problems described above is a maximization over set of all positive stopping times $\tau$.

Maximization over the same set!
Three abstract indices

Three abstract optimization problems
Suppose there is an \textit{abstract index set} $U$, and $A = \{a_u\}$ and $B = \{b_u\}$ be two sets indexed by the elements of $U$. Suppose that an assumption $U$ holds,

\begin{align*}
a_u \leq a < \infty, \quad 0 < b \leq b_u \leq 1 \quad (U)
\end{align*}

Problem 1. Restart problem  (from Katehakis- Veinott index) Find solution(s) of the equation

\begin{align*}
h = \sup_{u \in U} [a_u + (1 - b_u)h], \quad \text{i.e.} \quad h = H(h), \quad (*)
\end{align*}

where $H(k) = \sup_{u \in U} [a_u + (1 - b_u)k]$. 

$h = \text{Abstract KV index}$
There are two equivalent interpretations of this problem.

There is a set of "buttons" $u \in U$. A DM can select one of them and push. She obtains a reward $a_u$ and according to the first interpretation with probability $b_u$ the game is terminated, and with complimentary probability $1 - b_u$ she can select any button again. Her goal is to maximize the total (undiscounted) reward.

According to the second interpretation the game is continued sequentially and $1 - b_u$ is not the probability but a discount factor applied to the future rewards. It can be easily proved that in both cases its value satisfies the equation above.
Our second optimization problem is

**Problem 2. Ratio (cycle) problem**

Find $\alpha$

$$\alpha = \sup_{u \in U} \frac{a_u}{b_u} \quad \text{(Gittins index)} \quad (***)$$

The interpretation of this problem is straightforward: a DM wants to maximize the ratio of the one step reward per "chance of termination".

$\alpha = \textbf{Abstract Gittins index}$
Problem 3. A parametric family of Retirement problems

Find \( w \), (abstract Whittle index) defined as follows: given parameter \( k, -\infty < k < \infty \), let \( \nu(k) = \max(k, H(k)) \), where

\[
H(k) = \sup_{u \in U} [a_u + (1 - b_u)k].
\]

\[
w = \inf \{k : \nu(k) = k\}. \quad (***)
\]

\[
h = \sup_{u \in U} [a_u + (1 - b_u)h]. \quad (*)
\]

\[
\alpha = \sup_{u \in U} \frac{a_u}{b_u} \quad (**) \]

Theorem 3

a) The solution \( h \) of the equation (*) exists, is unique and finite;
b) \( h = \alpha = w \); c) the optimal index \( u \) (or an optimizing sequence \( u_n \)) for any of three problems is the optimal index (or an optimizing sequence) for two other problems.

The proof is elementary. Function \( H(k) \) is nondecreasing, continuous, and convex.
Theorem 2 from Theorem 3: $U = \{u\} = \{\text{set of all Markov moments } \tau > 0\}$, $a_u = R^\tau(x) = E_x \sum_{n=0}^{\tau-1} c(Z_n)$, the total expected reward till moment $\tau$, and $b_u = Q^\tau(x) = P_x(Z_\tau = e)$, the probability of termination on $[0, \tau)$.

**Remark**! Theorem 3: Only equivalence, not how to solve!

**Problem 4.** Suppose that a DM has to solve the optimization problem similar to Problem 1 with sequential selection of buttons with only one distinction - every button can be used *at most once*.

The Mitten’s result (1960) essentially can be described as

---

**Theorem 4**

*Suppose that there is a sequence of indices $u_n$ such that after the relabeling of indices in this sequence, we have $\alpha_1 = \frac{a_1}{b_2} \geq \alpha_2 = \frac{a_2}{b_2} \geq ... \geq \frac{a_u}{b_u}$ for each $u \in U$ not in this sequence. Then to push buttons in the order $1, 2, ..$ is an optimal strategy.*
(joint with S. Steinberg)
A general CQR model is specified by a tuple 
\( M = (X, B, P, A(x), c, q, r_i(x)) \), where \( X \) is a countable state space, 
\( B = \{s_1, ..., s_m\} \) is a subset of a state space \( X \), 
\( P = \{p(x, y)\} \) a stochastic matrix
set of available actions \( A(x) = (\text{continue, quit, restart to } s_i) \)
reward function \( r(x|a) \) is specified by particular functions \( c(x), q(x) \) and 
\( r_i(x), i = 1, 2, ..., m. \)
If an action \( c \), ”continue” is selected then \( r(x|c) = c(x) \) and transition to a 
new state occurs according to transition probabilities \( p(x, y) \), if an action \( q \), 
”quit” is selected then \( r(x|q) = q(x) \) and transition to an absorbing state \( e \) 
occurs with probability one, if an action \( r_i \), ”restart to state \( s_i \)” is selected 
then \( r(x|r_i) = r_i(x) \) and transition to a state \( s_i \) occurs with probability one.
Idea: Sonin (2008), St. & Pr. Let. no quit, restart with zero fee;
a family of Whittle OS models $M(k) = (X, P, c(x), g(x|k), \beta(x))$, terminal reward function $g(x|k) = k$.

Consider function $G(x|k) = g(x|k) - [c(x) + Pg(x|k)]$. Linear function. Move $k$ from large values to the left. Eliminate and Eliminate, order of $n^3$.

Now, more complicated: $g(x|k) = \max(q(x), (r(x) + k))$, $G(x|k)$ is only piecewise linear!
Eliminate and Insert. Eliminate and Insert...

**Theorem 5**

For $m = 1$ and finite $X$ there is an algorithm solving this problem in a finite number of steps; order of $n^4$.  

---

Isaac M. Sonin 
Optimal Stopping of Markov chain, Gittins Index and Related Optimization Problems 

New York, Columbia University, September 2011 

24/25
Open Problems.

- multiple restart problems
- multidimensional equivalent abstract optimization problems
  \[ h = \sup_u [a_u + B_u h], \text{ vectors and matrices or...} \]
- explanation of world financial crisis

Thank you for your attention!