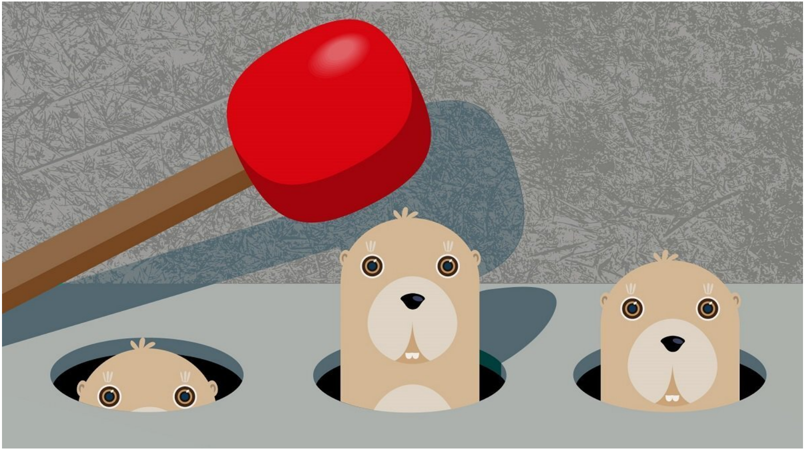
Restless bandits with controlled restarts: Indexability and computation of Whittle index

Nima Akbarzadeh, Aditya Mahajan

McGill University, Electrical and Computer Engineering Department

Dec. 13, 2019

Whack a Mole



Applications

Applications: queueing, channel scheduling, machine maintenance and clinical care.

1. A repairman is responsible for **maintaining** several machines. Each machine stochastically **deteriorates**. There is a state-dependent **cost** associated with running and repairing the machine. He can repair one machine at a time.
2. Scheduling **multiple data queues** over a shared communication channels, there is a **cost** associated with holding packets or transmitting it. A fixed number of data queues can be selected at a time.

The machine/queue restarts upon being repaired/selected.

Goal: Find a optimal/near-optimal policy to optimize scheduling!

Applications

Applications: queueing, channel scheduling, machine maintenance and clinical care.

1. A repairman is responsible for **maintaining** several machines. Each machine stochastically **deteriorates**. There is a state-dependent **cost** associated with running and repairing the machine. He can repair one machine at a time.
2. Scheduling **multiple data queues** over a shared communication channels, there is a **cost** associated with holding packets or transmitting it. A fixed number of data queues can be selected at a time.

The machine/queue restarts upon being repaired/selected.
Goal: Find a optimal/near-optimal policy to optimize scheduling!

Model

- n available arms (controlled Markov processes), $\mathcal{N} = \{1, \ldots, n\}$.
- m arms have to be selected. $(m < n)$
- State space of each arm $\mathcal{X}^i$, $i \in \mathcal{N}$
- Action space for each arm $\{0, 1\}$
- Passive action: $a_t^i = 0 \rightarrow$ Markov chain matrix P_{xy}^i
- Active action: $a_t^i = 1 \rightarrow$ Reset PMF Q_y^i
- Cost: $c^i(x_t^i, a_t^i)$

Objective

Problem

Given the discount factor β, the total number n of arms, the number m of active arms, the state space $\{\mathcal{X}^i\}_{i\in\mathcal{N}}$, the transition matrices $\{P^i\}_{i\in\mathcal{N}}$, the reset pmfs $\{Q^i\}_{i\in\mathcal{N}}$, and the cost functions $\{c^i(\cdot,\cdot)\}_{i\in\mathcal{N}}$,
choose a time-homogeneous Markov policy $\boldsymbol{g}$,

$$\boldsymbol{A}_t = \boldsymbol{g}(\boldsymbol{X}_t) \text{ such that } \sum_{i\in\mathcal{N}} A_t^i = m$$

that minimizes

$$J(\boldsymbol{g}) := (1-\beta)\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \sum_{i\in\mathcal{N}} c^i(X_t^i, A_t^i)\right].$$

Challenge & Solution

Challenge: The dynamic program suffers from curse of dimensionality! The size of the state space is $|\mathcal{X}|^n$.
Example: 100 machines with 3 states each results in a system with $3^{100} \approx 5.15 \times 10^{47}$ states!

Solution: Index-based heuristic policy (Whittle index [1988])
Drawback: Suboptimal!
Advantage: Problem decomposition $\Rightarrow$ 100 problems with 3 states.

Challenge & Solution

Challenge: The dynamic program suffers from curse of dimensionality! The size of the state space is $|\mathcal{X}|^n$.
Example: 100 machines with 3 states each results in a system with $3^{100} \approx 5.15 \times 10^{47}$ states!

Solution: Index-based heuristic policy (Whittle index [1988])
Drawback: Suboptimal!
Advantage: Problem decomposition $\Rightarrow$ 100 problems with 3 states.

Whittle Index policy

- Whittle index heuristic provides a dynamic index for each arm and select the arm with the smallest index at each time.
- Whittle index exists if indexability condition is satisfied for all arms.
- Whittle index policy performs close-to-optimal for many applications in the state-of-arts works.
- There is no general framework to check indexability and correspondingly, obtain the Whittle indices.

Objectives:

- Prove our problem is **indexable**.
- Provide a closed-form solution for the **Whittle index**.

Whittle Index policy

- Whittle index heuristic provides a dynamic index for each arm and select the arm with the smallest index at each time.
- Whittle index exists if indexability condition is satisfied for all arms.
- Whittle index policy performs close-to-optimal for many applications in the state-of-arts works.
- There is no general framework to check indexability and correspondingly, obtain the Whittle indices.

Objectives:

- Prove our problem is **indexable**.
- Provide a closed-form solution for the **Whittle index**.

Problem Decomposition

Define

$$c_\lambda(x_t^i, a_t^i) := c^i(x^i, a_t^i) + \lambda a_t^i,\ \ a_t^i \in \{0,1\}$$

for arm i.

Problem

Given an arm $i \in \mathcal{N}$, discount factor β, the state space $\mathcal{X}^i$, the transition probability matrix P^i, the reset probability mass function Q^i, the cost function $c^i(\cdot,\cdot)$ and the penalty $\lambda \in \mathbb{R}$,
choose a policy $g^i : \mathcal{X}^i \to \{0,1\}$ *to* **minimize**

$$J^i(g^i) := (1-\beta)\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t c_\lambda^i(X_t^i, A_t^i)\right].$$

Dynamic Programming

Theorem

Let $V_\lambda^i : \mathcal{X}^i \to \mathbb{R}$ *be the unique fixed point of the following:*

$$V_\lambda^i(x) = \min\big\{H_\lambda^i(x,0), H_\lambda^i(x,1)\big\}, \ \forall x \in \mathcal{X}^i.$$

where

$$H_\lambda^i(x,0) = (1-\beta)c^i(x,0) + \beta \sum_{y \in \mathcal{X}^i} P_{xy}^i V_\lambda^i(y),$$

$$H_\lambda^i(x,1) = (1-\beta)\left(c^i(x,1) + \lambda\right) + \beta \sum_{y \in \mathcal{X}^i} Q_y^i V_\lambda^i(y).$$

Let $g_\lambda^i(x)$ *denote the minimizer of the right hand side. Then,* g_λ^i *is optimal for arm* i*.*

Indexability

Let passive set for arm i be

$$\Pi_\lambda^i := \{x^i \in \mathcal{X}^i : g_\lambda^i(x) = 0\}.$$

Definition (Indexability)

For any $\lambda_1, \lambda_2 \in \mathbb{R}$ arm i is indexable if

$$\lambda_1 < \lambda_2 \implies \Pi_{\lambda_1}^i \subseteq \Pi_{\lambda_2}^i.$$

Definition (Whittle index)

The Whittle index of state x of arm i is defined as

$$w^i(x) = \inf\left\{\lambda \in \mathbb{R} : x \in \Pi_\lambda^i\right\}.$$

Indexability Proof Sketch

Theorem

Each arm is indexable.

Lemma

$$\Pi_\lambda = \left\{ x \in \mathcal{X} : (1-\beta) \inf_\tau \frac{L(x,\tau) - c(x,1)}{1-\beta^\tau} < W_\lambda \right\}.$$

Lemma

$W_\lambda = \lambda + \beta \sum_{y \in \mathcal{X}} Q_y V_\lambda(y)$ *is increasing in* λ.

Whittle index

By definition,

$$w^i(x) = \inf \left\{ \lambda \in \mathbb{R} : (1-\beta) \inf_{\tau} \frac{L(x,\tau) - c(x,1)}{1-\beta^{\tau}} < \lambda + \beta \sum_{y \in \mathcal{X}^i} Q^i_y V^i_\lambda(y) \right\}.$$

Challenge: Obtaining a closed form solution for Whittle index is inefficient.
Solution: To provide a closed-form solution we consider threshold-based policies.

Threshold Policies

The optimal policy for each subproblem is a threshold-based policy, i.e.,

$$g^{(k)}(x) := \begin{cases} 0, & \text{if } x < k \\ 1, & \text{otherwise.} \end{cases}$$

$$C_\lambda^{(k)} := (1-\beta)\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t c_\lambda(X_t, g^{(k)}(X_t)) \;\middle|\; X_0 \sim Q\right] = D^{(k)} + \lambda N^{(k)}.$$

where

$$D^{(k)} := (1-\beta)\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t c(X_t, g^{(k)}(X_t)) \;\middle|\; X_0 \sim Q\right],$$

$$N^{(k)} := (1-\beta)\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t g^{(k)}(X_t) \;\middle|\; X_0 \sim Q\right].$$

Computation of $D^{(k)}$ and $N^{(k)}$

Let

$$L^{(k)} := \mathbb{E}\left[\sum_{t=0}^{\tau_k - 1} \beta^t c(X_t, 0) + \beta^{\tau_k} c(X_{\tau_k}, 1) \,\middle|\, X_0 \sim Q \right]$$

$$M^{(k)} := \mathbb{E}\left[\sum_{t=0}^{\tau_k} \beta^t \,\middle|\, X_0 \sim Q \right].$$

Theorem

For all threshold k,

$$D^{(k)} = \frac{L^{(k)}}{M^{(k)}} \quad \text{and} \quad N^{(k)} = \frac{1}{\beta M^{(k)}} - \frac{1-\beta}{\beta}.$$

Property

Lemma

$k_\lambda := \arg\min_{k\in\mathcal{X}} C_\lambda^{(k)}$ is increasing in λ.

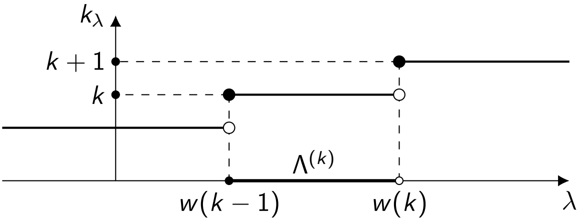


Figure: k_λ as a function of λ.

Whittle Index

Theorem

The Whittle index for threshold-policies at state $k \in \mathcal{X}$ is

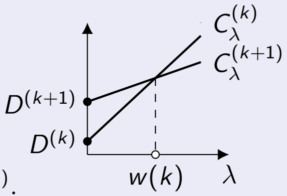
$$w(k) = \frac{D^{(k+1)} - D^{(k)}}{N^{(k)} - N^{(k+1)}}.$$

Proof.

Key Ideas:

- $C_\lambda^{(k)}$ is continuous in λ.
- $C_{w(k)}^{(k)} = C_{w(k)}^{(k+1)}$, i.e.,

$$D^{(k)} + w(k)N^{(k)} = D^{(k+1)} + w(k)N^{(k+1)}.$$



□

Whittle Index policy

- Compute Whittle indices offline.
- At each time instance, observe the state of each arm and select the arm with the **lowest** Whittle index.

Experiment Setup

- **Deterministic restart**: $Q = [1, 0, \ldots, 0]$
- $c(x, 0) = (x - 1)^2$ and $c(x, 1) = 0.5(|\mathcal{X}| - 1)^2$, $\beta = 0.9$
- We consider structured and randomly generated stochastic monotone matrices for P.
- **Monte-Carlo simulations**: 5000 iterations with 250 time steps in each one.

Experiments (1) & (2)

Comparison with **Optimal Policy** for small-scale models:

$$\alpha_{\mathrm{OPT}} = \frac{J(\mathrm{OPT})}{J(\mathrm{WIP})} \times 100$$

For $|\mathcal{X}| = 5,\ n = 5,\ m \in \{1, 2\} \rightarrow \alpha_{\mathrm{OPT}} \in [95.5\% - 100\%]$.

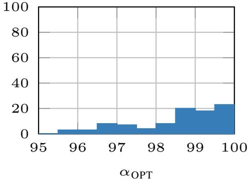


Figure: 100 randomly generated stochastic monotone matrices with $m = 1$.

Experiments (3) & (4)

Comparison with **Myopic Policy** for large-scale models:

$$\varepsilon_{\text{MYP}} = \left(\frac{J(\text{MYP}) - J(\text{WIP})}{J(\text{MYP})} \right) \times 100.$$

For $|\mathcal{X}| = 25,\ n \in \{25, 50, 75\},\ m \in \{1, 2, 5\}$
$\rightarrow \varepsilon_{\text{MYP}} \in [0\% - 12\%]$.

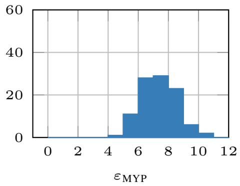


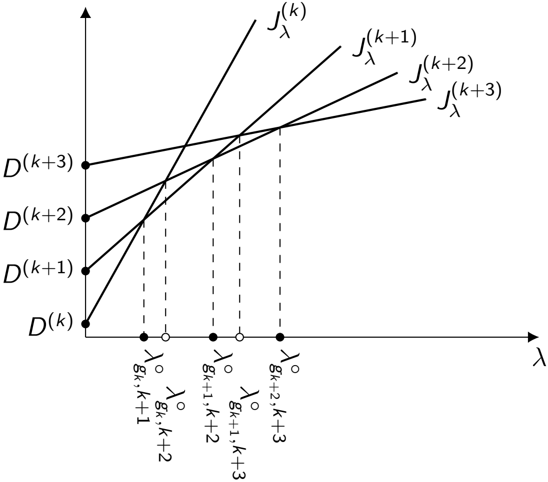
Figure: 100 randomly generated stochastic monotone matrices with $n = 75,\ m = 2$.

Conclusion

- A model for restless bandit with controlled restarts.
- An indexable model.
- A closed form expression to compute the Whittle indices when the optimal policy is threshold-based.
- Numerical experiments shows the Whittle index policy performs very close to the optimal policy and better than a myopic policy.

Thank you!

Q&A



$J_\lambda^{(k)}$
$J_\lambda^{(k+1)}$
$J_\lambda^{(k+2)}$
$J_\lambda^{(k+3)}$
$D^{(k+3)}$
$D^{(k+2)}$
$D^{(k+1)}$
$D^{(k)}$
λ
$\lambda^\circ_{g_k,k+1}$
$\lambda^\circ_{g_k,k+2}$
$\lambda^\circ_{g_{k+1},k+2}$
$\lambda^\circ_{g_{k+1},k+3}$
$\lambda^\circ_{g_{k+2},k+3}$