Sequential Selection of Projects

Kemal Gürsoy

Rutgers University, Department of MSIS, New Jersey, USA

Fusion Fest

October 11, 2014

RUTGERS

Outline

1. **Introduction**
   - Model
2. **Necessary Knowledge**
   - Sequential Statistics
   - Multi-Armed Bandits
3. **Conclusion**
   - Work Done and Future Work

RUTGERS

Competing projects.

Assumptions

- Each project i has a positive reward R_i, upon its completion.
- The completion time of each project i is a positive and conditionally independent random variable $\tau_i \sim F_i(x_i, t_i)$, based on the state, x_i, and the activation time, t_i.
- The expected reward of a project i depends upon its completion time, $E[R_i e^{-\alpha\tau_i} | x_i, t_i]$.
- Where $\alpha \in (0, 1)$ is the time-discount factor for all projects.

Construction of the *Selection Policy.*

Construction

Let there be a set of projects such that for a pair i, j

- A selection policy orders activation times of these projects, $E[R_i e^{-\alpha\tau_i} + R_j e^{-\alpha(\tau_i+\tau_j)}] > E[R_j e^{-\alpha\tau_j} + R_i e^{-\alpha(\tau_j+\tau_i)}]$
- Due to the linearity property of the expectation operator: $E[R_i e^{-\alpha\tau_i}] + E[R_j e^{-\alpha(\tau_i+\tau_j)}] > E[R_j e^{-\alpha\tau_j}] + E[R_i e^{-\alpha(\tau_j+\tau_i)}]$
- By the independence assumption of the completion times: $E[R_i e^{-\alpha\tau_i}] + E[R_j e^{-\alpha\tau_i}]E[R_j e^{-\alpha\tau_j}] >$ $E[R_j e^{-\alpha\tau_j}] + E[R_i e^{-\alpha\tau_j}]E[R_i e^{-\alpha\tau_i}]$
- By organizing similar terms: $\frac{E[R_i e^{-\alpha\tau_i}]}{E[1-e^{-\alpha\tau_i}]} > \frac{E[R_j e^{-\alpha\tau_j}]}{E[1-e^{-\alpha\tau_j}]}$.

The optimal activation policy

An ordering policy selects projects based on the diminishing values of $\frac{E[R_i e^{-\alpha\tau_i}]}{E[1-e^{-\alpha\tau_i}]}$. Let $g_i = \frac{E[R_i e^{-\alpha\tau_i}]}{E[1-e^{-\alpha\tau_i}]}$ be an activation index for the project i, such that $g_{[1]}$ is the maximum and $g_{[N]}$ is the minimum of g_i values.

Theorem

Optimal activation policy is identified by the ordering of $g_{[i]}$s; $g_{[1]} > g_{[2]} > \cdots > g_{[N-1]} > g_{[N]}$.

Sketch of the Proof.

Activate an inferior value project first, this will delay the activation of the superior value project. This is not the best discounted expected total reward. □

Sequentially selecting subsets of projects.

There is an optimal policy for activating an ensemble of projects.

1. Compute g_i and order all the projects with the decreasing values of g_i. This ordering identifies an index set for an optimal activation policy.
2. Fix a subset cardinality, say k, of projects to be activated simultaneously.
3. Select the first k number of projects and activate them.
4. Continue activating the ensemble of k projects, based on the remaining elements of the ordered list, until all the projects are completed.

Proof is by deduction.

RUTGERS

Sequential experimentation

In the sequential design of experiments, the size of the samples are not fixed in advance, but are functions of observations.

A brief timeline of the sequential experimentation:

- Statistical quality control of Dodge and Romig (1929)
- Sampling design of Mahalonobis (1940)
- Sequential analysis of Wald (1947)
- Sequential design of experiments by Robbins (1952)

Multi-armed bandit problem.

The multi-armed bandit problem is a statistical model for the adaptive control problems, formulated by Herbert E. Robbins (1952).

Some important contributions are works of Karlin (1956), Chernoff (1965), Gittins and Jones (1974), Whittle (1980).

- The multi-armed bandits are Bernoulli reward processes.
- These semi-Markov decision processes are independent.
- Bandits represent generalized projects.

Computations.

- Gittins and Jones designed an index to identify the activation order of the multi-armed bandits (1972), by assuming a preemptive scenario.

  Gittins Index: $\nu_i(x_{t_0}) = \sup_\tau \frac{E[\sum_{t=t_0}^{\tau-1} \alpha^t r(x_t)]}{E[\sum_{t=t_0}^{\tau-1} \alpha^t]}$.

  Where $r(x_t)$ is the reward provided by the ith bandit at its state x_t, and τ is its stopping time. Gittins index points at the project to be activated, and also for how long it should be activated.

- Katehakis and Veinott (1987) constructed an efficient computation for the Gittins indices, based on the restart in the reward state formulation.

The modified problem.

Work done in the generalization.

- Simultaneous projects.
- Influential projects.

Future direction.

- Dependent Markov decision processes.

Dear Paul, I wish you the best.

References I

J. Gittins, K. Glazebrook, R. Weber.
Multi-Armed Bandit Allocation Indices.
Wiley, 2011.

H.E. Robbins.
Some aspects of the sequential design of experiments.
Bulletin of The American Mathematical Society, Vol.58(5): 527–535, 1952.

M.N. Katehakis, A.F. Veinott Jr.
The multiarmed bandit problem: Decomposition and computation.
Mathematics of Operations Research, 12(2): 262–268, 1987.