

Lévy bandits under Poissonian decision times

José-Luis Pérez and Kazutoshi Yamazaki

Abstract We consider a version of the continuous-time multi-armed bandit problem where decision opportunities arrive at Poisson arrival times, and study its Gittins index policy. When driven by spectrally one-sided Lévy processes, the Gittins index can be written explicitly in terms of the scale function, and is shown to converge to that in the classical Lévy bandit of Kaspi and Mandelbaum [8].

Key words: multi-armed bandits, Gittins index, Lévy processes, stochastic control
AMS 2020 Subject Classifications: 60G51, 93E20, 90B36

1 Introduction

The multi-armed bandit (MAB) problem is a stochastic dynamic optimization problem where at each decision time, one of J alternatives (called *arms*) is selected and a reward is then collected. In its infinite-time horizon formulation, the objective is to derive a policy (strategy) that maximizes the expected total discounted rewards. An important assumption of the MAB is that the state of each arm changes only when it is selected and the evolution of each arm is independent of others. By exploiting these features, the existing results have shown the optimality of the so-called *Gittins index* policy [6], which selects at each time the arm with the highest Gittins index. The Gittins index is defined in terms of an optimal stopping problem and can be computed separately for each arm. The state of each arm is usually assumed to evolve in a Markovian fashion, and thus the MAB can be seen as an instance of

José-Luis Pérez

Department of Probability and Statistics, Centro de Investigación en Matemáticas A.C. Calle Jalisco s/n. C.P. 36240, Guanajuato, Mexico, e-mail: jluis.garmendia@cimat.mx

Kazutoshi Yamazaki

School of Mathematics and Physics, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia e-mail: k.yamazaki@uq.edu.au

the Markov decision process (MDP) problem (see [22]). This optimality result is important in the MDP research in that the original J -dimensional MDP problem can be reduced to J independent 1-dimensional MDPs, providing a way to overcome the *curse of dimensionality*.

The continuous-time MAB theory has been developed in the 1990s by, for example, El Karoui and Karatzas [7] and Kaspi and Mandelbaum [8, 9]. For the setting where an arm is continuously selected, advanced stochastic analysis techniques, in particular local time analysis, are required to precisely model the problem. However, main results in the discrete-time model continue to hold in the continuous-time model. The Gittins index can be defined in a similar way, and it is optimal to choose the arm with the highest Gittins index at all times. The advantage of studying a continuous-time model is that the Gittins index can be sometimes obtained in an explicit form while numerical approaches are usually required in the discrete time case. In particular, Kaspi and Mandelbaum [8] considered the model driven by one-dimensional Lévy processes and expressed the Gittins index concisely using the Wiener-Hopf factorization and excursion theory of Lévy processes.

In this paper, we revisit the MAB problem, in particular the Lévy model of [8], and study a version where decision opportunities arrive only at independent Poisson arrival times. This problem is at the interface between the discrete-time and continuous-time models. The state of each arm evolves continuously whereas decision times are discrete. We consider two variations of the problem and study the corresponding Gittins index policy for each. We introduce cutting-edge results of the fluctuation theory of Lévy processes (e.g. [3, 11]), which had not been available when the results of [8] were developed.

The Poissonian observation model has recently drawn much attention in the stochastic control literature (see, e.g., [2, 5, 18, 19, 21, 24]), but it has not been considered in the MAB problem. With the assumption that decision opportunities arrive at exponential times, an analytical approach is still possible thanks to recently developed fluctuation-theory-based methods such as [1, 15, 20]. The Gittins index can be written in terms of the so-called *scale function* when the Lévy process has only one-sided jumps (spectrally one-sided). We further investigate its connection with [8] by showing the convergence of the Gittins index to that obtained in [8].

The rest of the paper is organized as follows. In Section 2, we review the classical discrete-time and continuous-time models of the MAB problem, as well as the Lévy model of [8]. In Section 3, we introduce the Poissonian decision time models and propose our Gittins index policy. In Section 4, we review the fluctuation theory of spectrally one-sided Lévy processes and the scale function, and then obtain the Gittins index explicitly in terms of the scale function. In Section 5, we show the convergence of our Gittins index to that in [8]. We conclude the paper in Section 6.

2 Preliminaries

2.1 Discrete-time Markovian bandits

A (multi-armed) Markovian bandit consists of $J \geq 1$ arms. For each arm labeled $j \in \mathcal{J} := \{1, \dots, J\}$, we use an E_j -valued discrete-time Markov process $X^j = (X^j(k))_{k \geq 0}$ defined on a probability space $(\Omega^j, \mathcal{F}^j, \mathbb{P}^j)$ to model the state dynamics of arm j . These J processes are assumed mutually independent. At each time $k \geq 0$, one of the J arms is selected. A state-dependent reward is then collected and the state of the arm changes. On the other hand, the states of all the other arms remain the same.

A policy (strategy) $\pi = (\pi(k))_{k \geq 0}$ models the arm to be selected at each time $k \geq 0$. It is adapted in the sense that $\pi(k)$ must be determined based only on the information collected until time k . Let $\sigma_j^\pi(k)$ be the number of times arm $j \in \mathcal{J}$ has been selected before time $k \geq 0$ under policy π . At time k , immediately before making the decision, the state of the arms are $(X^j(\sigma_j^\pi(k)), j \in \mathcal{J})$. Using these, the decision maker needs to determine $\pi(k)$ and the reward from arm $\pi(k)$ is collected. The counters $(\sigma_j^\pi; j \in \mathcal{J})$ are updated as follows: $\sigma_j^\pi(k+1) = \sigma_j^\pi(k) + 1$ if $j = \pi(k)$ while $\sigma_j^\pi(k+1) = \sigma_j^\pi(k)$ if $j \neq \pi(k)$.

The objective of the MAB problem is to obtain an optimal policy that maximizes the total expected reward over an infinite-time horizon. Given a discount factor $0 < \beta < 1$ and a reward function $R^j : E_j \rightarrow \mathbb{R}$ for each $j \in \mathcal{J}$, the problem is to compute the value function

$$v(\mathbf{x}) = \sup_{\pi} \mathbb{E}_{\mathbf{x}} \left[\sum_{k=0}^{\infty} \beta^k R^{\pi(k)}(X^{\pi(k)}(\sigma_{\pi(k)}^\pi(k))) \right], \quad (1)$$

for $\mathbf{x} = (x^1, \dots, x^J) \in E_1 \times \dots \times E_J$ and obtain an optimal strategy π^* . Here, $\mathbb{E}_{\mathbf{x}}$ is the conditional expectation when $X^j(0) = x^j$ for $j \in \mathcal{J}$.

In this problem, there exists a simple policy, known as the Gittins index policy, that works optimally in every instance of the MAB problem. It is optimal to choose the arm with the highest Gittins index, i.e. to select, at each $k \geq 0$,

$$\pi^*(k) = \arg \max_{j \in \mathcal{J}} \Gamma_j(X^j(\sigma_j^{\pi^*}(k))),$$

where the Gittins index of arm j at state x is a “normalized maximal reward”

$$\Gamma_j(x) := \sup_{\tau \geq 1} \frac{\mathbb{E}_x^j \left[\sum_{k=0}^{\tau-1} \beta^k R^j(X^j(k)) \right]}{\mathbb{E}_x^j \left[\sum_{k=0}^{\tau-1} \beta^k \right]}, \quad (2)$$

which can be computed without any knowledge about the other arms. Here, τ is chosen over all stopping times greater than or equal to 1 with respect to the filtration generated by $(X^j(k))_{k \geq 0}$, and \mathbb{E}_x^j is the expectation operator of the law of \mathbb{P}_x^j under which $X^j(0) = x$ assuming arm j is always selected.

The Gittins index (2) can be derived by computing the unique value of γ such that the value of the optimal stopping problem

$$\sup_{\tau \geq 1} \mathbb{E}_x \left[\sum_{k=0}^{\tau-1} \beta^k (R^j(X^j(k)) - \gamma) \right] \quad (3)$$

becomes zero; see [9, Eqn. (20)].

2.2 Continuous-time bandits

In the continuous-time model, the state process $X^j = (X^j(t))_{t \geq 0}$, for $j \in \mathcal{J}$, is an E_j -valued continuous-time process defined on a probability space $(\Omega^j, \mathcal{F}^j, \mathbb{P}^j)$. Again, these processes are assumed mutually independent.

The evolution of the arms is given by

$$\mathbf{X} := (\mathbf{X}(\mathbf{s}) = (X^1(s_1), \dots, X^J(s_J)), \mathbf{s} = (s_1, \dots, s_J) \geq \mathbf{0}),$$

which is adapted to the multiparameter filtration $(\mathcal{F}(\mathbf{s}) = \mathcal{F}^1(s_1) \vee \dots \vee \mathcal{F}^J(s_J))_{\mathbf{s} \geq \mathbf{0}}$ where $(\mathcal{F}^j(t))_{t \geq 0}$ is the natural filtration generated by X^j for $j \in \mathcal{J}$. For more details, we refer the reader to [8].

In the continuous-time model, a strategy is given by a J -dimensional process

$$\mathbf{T} = (\mathbf{T}(t) = (T^1(t), \dots, T^J(t)); t \geq 0)$$

where $T^j(t)$ is the amount of time allocated to arm j until t . Naturally, $\mathbf{T}(t)$ is nondecreasing, and $T^1(t) + \dots + T^J(t) = t$ for all $t \geq 0$. A strategy has to be adapted and hence

$$\{T^1(t) \leq s_1, \dots, T^J(t) \leq s_J\} \in \mathcal{F}(\mathbf{s})$$

for all $t \geq 0$ and $\mathbf{s} = (s_1, \dots, s_J) \in [0, \infty)^J$.

Given a discount factor $q > 0$ and a reward function $r^j : E_j \rightarrow \mathbb{R}$, $j \in \mathcal{J}$, the objective is to maximize the total discounted expected reward

$$v(\mathbf{x}) = \sup_{\mathbf{T}} \mathbb{E}_{\mathbf{x}} \left[\int_0^\infty e^{-qt} \sum_{j \in \mathcal{J}} r^j(X^j(T^j(t))) dT^j(t) \right]$$

over all allocation strategies \mathbf{T} . Regarding the reward function, for the problem to be well-defined, it is assumed that

$$\mathbb{E}_x^j \left[\int_0^\infty e^{-qt} |r^j(X^j(t))| dt \right] < \infty, \quad j \in \mathcal{J}, x \in E_j \quad (4)$$

where again \mathbb{E}_x^j is the expectation operator of the law \mathbb{P}_x^j of X^j under which $X^j(0) = x$ assuming arm j is always selected.

The Gittins index in the continuous-time model is analogous to (2) where the summations are replaced by integrals:

$$\Gamma_j(x) := \sup_{\tau > 0} \frac{\mathbb{E}_x^j \left[\int_0^\tau e^{-qt} r^j(X^j(t)) dt \right]}{\mathbb{E}_x^j \left[\int_0^\tau e^{-qt} dt \right]}, \quad j \in \mathcal{J}, x \in E_j. \quad (5)$$

Similar to the discrete-time counterpart, it is optimal to choose the arm with the highest Gittins index, although some technical details are needed to deal with the times at which multiple Gittins indices coincide.

2.3 Lévy bandits

As a special case of the continuous-time model, in Kaspi and Mandelbaum [8], an explicit expression of the Gittins index (5) was obtained under the following assumption: for each $j \in \mathcal{J}$,

1. the state process X^j is an \mathbb{R} -valued Lévy process ($E_j = \mathbb{R}$);
2. the reward function r^j is increasing and continuous.

Under these assumptions, the Gittins index (5) for arm $j \in \mathcal{J}$ when its state is $x \in \mathbb{R}$ can be written concisely by

$$\Gamma_j(x) = \int_{[0, \infty)} r^j(x+y) \mu^j(dy), \quad x \in \mathbb{R},$$

where μ^j is the probability measure determined by

$$\int_{[0, \infty)} e^{-\theta y} \mu^j(dy) = \frac{\bar{\varphi}^j(q, 0)}{\bar{\varphi}^j(q, \theta)}, \quad \theta \geq 0, \quad (6)$$

in terms of the Laplace exponent $\bar{\varphi}^j(q, \theta)$ of the inverse local time at maximum and the ascending ladder height process $(L^{-1}(t), H(t))_{t \geq 0}$ (see Section 5.1). The above characterization in terms of (6) is given in (4.5) of [8], which is derived from an alternative expression (3.3) of [8] via the Wiener-Hopf factorization.

The optimal value function can be written in terms of the Gittins index function. We have

$$v(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} \left[\int_0^\infty e^{-qt} \bigvee_{j \in \mathcal{J}} \Gamma_j[\underline{X}^j(T^{*,j}(t))] dt \right]$$

where $\underline{X}^j(t) := \inf_{0 \leq u \leq t} X^j(u)$ is the running infimum process and $(T^{*,j}; j \in \mathcal{J})$ is the allocation time under the Gittins index policy.

3 Periodic observation models

We now consider a version of the continuous-time model where decision times arrive at exponential times. Let $N = (N(t))_{t \geq 0}$ be a Poisson process with rate $\lambda > 0$, independent of X^j for all $j \in \mathcal{J}$, with arrival times $\mathcal{T} := \{T_1, T_2, \dots\}$ (hence $T_{k+1} - T_k, k \geq 0$, is an independent exponential time with mean λ^{-1} where $T_0 := 0$ for convenience). We assume the decision times are given by $\mathcal{T} \cup \{0\}$. For each $T_k, k \geq 0$, one arm is selected. If arm j is selected at T_k , its state changes as a Lévy process according to the law \mathbb{P}^j until the next decision time T_{k+1} , while the states of the other arms remain the same during the period.

We consider two variations of the problem when X^j is a one-dimensional Lévy process for each $j \in \mathcal{J}$.

3.1 Problem 1

With a discount factor $q > 0$ and a reward function $R^j : \mathbb{R} \rightarrow \mathbb{R}$, we consider the maximization problem:

$$v(\mathbf{x}) = \sup_{\pi} \mathbb{E}_{\mathbf{x}} \left[\sum_{k=0}^{\infty} e^{-qT_k} R^{\pi(k)}(X^{\pi(k)}(\sigma_{\pi(k)}^{\pi}(k))) \right],$$

which can be seen as a version of the discrete-time model (1) where the deterministic decision times $\mathbb{N} \cup \{0\}$ are replaced by the Poisson arrival times (in addition to zero) $\mathcal{T} \cup \{0\}$. Here, analogously to (1), $\pi(k) \in \mathcal{J}$ denotes the arm selected at time T_k and

$$\sigma_j^{\pi}(k) := \sum_{i=0}^{k-1} (T_{i+1} - T_i) 1_{\{\pi(i)=j\}}$$

is the amount of times arm j has been active before T_k under policy π . The selection $\pi(k)$ must be determined based only on the information collected right before making decisions, i.e.

$$T_1, \dots, T_k \quad \text{and} \quad X^j(t), t \leq \sigma_j^{\pi}(k) \quad \text{for all } j \in \mathcal{J}. \quad (7)$$

Analogous to what is assumed in [8], we assume the following.

Assumption 3.1 For each $j \in \mathcal{J}$, we assume

1. the reward function R^j is strictly increasing,
2. We have $\mathbb{E}_x^j \left[\sum_{k=0}^{\infty} e^{-qT_k} |R^j(X^j(T_k))| \right] < \infty$ for $x \in \mathbb{R}$, where \mathbb{E}_x^j is the same as that in the continuous-time model in Section 2.2.

3.2 Problem 2

As a special case of Problem 1, we also consider a variant of the continuous-time model reviewed in Section 2.2, where the change of selection can be made only at $\mathcal{T} \cup \{0\}$. At each T_k , an arm is selected and it is locked in until T_{k+1} . The difference from Problem 1 is that the reward depends on the path between the decision times. This can be modeled as a maximization problem

$$v(\mathbf{x}) = \sup_{\pi} v_{\pi}(\mathbf{x}),$$

with

$$v_{\pi}(\mathbf{x}) := \mathbb{E}_{\mathbf{x}} \left[\int_0^{\infty} e^{-qt} \sum_{j \in \mathcal{J}} r^j (X^j(S_j^{\pi}(t))) dS_j^{\pi}(t) \right],$$

where, for each $j \in \mathcal{J}$, $r^j : \mathbb{R} \rightarrow \mathbb{R}$ and $S_j^{\pi}(t)$ is the amount of time at which j has been active until time t under π . For $T_k \leq t < T_{k+1}$, $S_{\pi(k)}^{\pi}(t) = \sigma_{\pi(k)}^{\pi}(k) + (t - T_k)$, and hence

$$v_{\pi}(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} \left[\sum_{k=0}^{\infty} \int_{T_k}^{T_{k+1}} e^{-qt} r^{\pi(k)} \left(X^{\pi(k)}(\sigma_{\pi(k)}^{\pi}(k) + (t - T_k)) \right) dt \right].$$

With $\mathcal{G}(T_k)$ all information observed until T_k under π (i.e. (7)) and e_{λ} an independent exponential random variable with mean λ^{-1} , noting that $\pi(k) \in \mathcal{G}(T_k)$ and because $T_{k+1} - T_k \sim e_{\lambda}$ and is independent of $\mathcal{G}(T_k)$, the strong Markov property gives

$$\begin{aligned} v_{\pi}(\mathbf{x}) &= \mathbb{E}_{\mathbf{x}} \left[\sum_{k=0}^{\infty} \mathbb{E}_{\mathbf{x}} \left(\int_{T_k}^{T_{k+1}} e^{-qt} r^{\pi(k)} \left(X^{\pi(k)}(\sigma_{\pi(k)}^{\pi}(k) + (t - T_k)) \right) dt \middle| \mathcal{G}(T_k) \right) \right] \\ &= \mathbb{E}_{\mathbf{x}} \left[\sum_{k=0}^{\infty} e^{-qT_k} \mathbb{E}_{\mathbf{x}} \left(\int_0^{e_{\lambda}} e^{-qt} r^{\pi(k)} \left(X^{\pi(k)}(\sigma_{\pi(k)}^{\pi}(k) + t) \right) dt \middle| \mathcal{G}(T_k) \right) \right] \\ &= \mathbb{E}_{\mathbf{x}} \left[\sum_{k=0}^{\infty} e^{-qT_k} R^{\pi(k)} \left(X^{\pi(k)}(\sigma_{\pi(k)}^{\pi}(k)) \right) \right], \end{aligned} \tag{8}$$

where

$$R^j(x) := \mathbb{E}_x^j \left[\int_0^{e_{\lambda}} e^{-qt} r^j (X^j(t)) dt \right] = \mathbb{E}_x^j \left[\int_0^{\infty} e^{-(q+\lambda)t} r^j (X^j(t)) dt \right], \quad j \in \mathcal{J} \text{ and } x \in \mathbb{R}. \tag{9}$$

Hence, this is a special case of Problem 1 with this choice of the reward function.

For Problem 2, we assume below, following what are assumed in [8].

Assumption 3.2 For each $j \in \mathcal{J}$, we assume

1. the reward function r^j is strictly increasing and continuous.
2. the integrability condition (4) is satisfied.

This assumption guarantees that (9) is finite and Assumption 3.1 is fulfilled.

3.3 Gittins index

We define the Gittins index for our problems, analogously to those in the classical case. For simplicity, we fix an arm $j \in \mathcal{J}$ and omit the index in an obvious manner (e.g. $R \equiv R^j$ and $X \equiv X^j$). Here, we follow similar steps as [9, Section 3.6] to derive our Gittins index.

Recall that the Gittins index in the discrete-time case (2) is derived by considering the optimal stopping problem (3) parameterized by γ . Thus, it is natural to consider an analogous index in the Poissonian decision time settings, by considering (3) with the deterministic decision times replaced by Poisson arrival times.

For each $\gamma \in \mathbb{R}$, consider the following auxiliary optimal stopping problem

$$v(x, \gamma) = \sup_{M \geq 1} \mathbb{E}_x \left[\sum_{k=0}^{M-1} e^{-qT_k} (R(X(T_k)) - \gamma) \right],$$

where M is selected from a set of stopping times greater than or equal to 1 with respect to the filtration $\mathbb{A} := (\mathcal{A}_k)_{k \geq 0}$ with $\mathcal{A}_k := \sigma(T_k, 1 \leq i \leq k, X(t), t \leq T_k)$, i.e. we stop at T_M for an \mathbb{A} -stopping time M .

Lemma 1. For $x \in \mathbb{R}$, the mapping $\gamma \mapsto v(x, \gamma)$ is strictly decreasing and convex.

Proof. For $\bar{\gamma} > \underline{\gamma}$ and an \mathbb{A} -stopping time $M \geq 1$,

$$\mathbb{E}_x \left[\sum_{k=0}^{M-1} e^{-qT_k} (R(X(T_k)) - \underline{\gamma}) \right] \geq (\bar{\gamma} - \underline{\gamma}) + \mathbb{E}_x \left[\sum_{k=0}^{M-1} e^{-qT_k} (R(X(T_k)) - \bar{\gamma}) \right].$$

After maximizing both sides over M , the decreasing property is immediate.

Regarding the convexity, for $\delta \in [0, 1]$, $\gamma_1, \gamma_2 \in \mathbb{R}$, and \mathbb{A} -stopping time M ,

$$\begin{aligned} v_M(x, \delta\gamma_1 + (1-\delta)\gamma_2) &:= \mathbb{E}_x \left[\sum_{k=0}^{M-1} e^{-qT_k} [R(X(T_k)) - (\delta\gamma_1 + (1-\delta)\gamma_2)] \right] \\ &= \delta \mathbb{E}_x \left[\sum_{k=0}^{M-1} e^{-qT_k} [R(X(T_k)) - \gamma_1] \right] + (1-\delta) \mathbb{E}_x \left[\sum_{k=0}^{M-1} e^{-qT_k} [R(X(T_k)) - \gamma_2] \right] \end{aligned}$$

where the decomposition makes sense from Assumption 3.1(2) together with the fact that $\mathbb{E}_x \left[\sum_{k=0}^{\infty} e^{-qT_k} \right] = \mathbb{E}_x [1 + \int_{(0, \infty)} e^{-qs} dN(s)] = (\lambda + q)/q < \infty$. This gives $v_M(x, \delta\gamma_1 + (1-\delta)\gamma_2) \leq \delta v(x, \gamma_1) + (1-\delta)v(x, \gamma_2)$ and hence $v(x, \delta\gamma_1 + (1-\delta)\gamma_2) \leq \delta v(x, \gamma_1) + (1-\delta)v(x, \gamma_2)$, showing the convexity.

Fix $x \in \mathbb{R}$. By Lemma 1 we have that the mapping $\gamma \mapsto v(x, \gamma)$ is continuous, and monotone convergence gives $\lim_{\gamma \downarrow -\infty} v(x, \gamma) = \infty$ and $\lim_{\gamma \uparrow \infty} v(x, \gamma) = -\infty$. Hence, $v(x, \cdot)$ has a unique root, which we denote by $\Gamma(x)$ such that

$$\begin{aligned} v(x, \gamma) > 0 &\iff \gamma < \Gamma(x), \\ v(x, \gamma) = 0 &\iff \gamma = \Gamma(x), \\ v(x, \gamma) < 0 &\iff \gamma > \Gamma(x). \end{aligned} \tag{10}$$

We call $\Gamma(x)$ the Gittins index for our problem.

By these obtained characteristics of the optimal stopping problem, following the same arguments in Section 3.6 of [9], we derive our Gittins index in the form analogous to (2).

By Snell's optimal-stopping theory (see, e.g., [23]) we have that the stopping rule that attains $v(x, \gamma)$ is given by

$$\begin{aligned} \tau(x, \gamma) &:= \inf\{m \geq 1 : v(X(T_m), \gamma) \leq 0\} \\ &= \inf\{m \geq 1 : \Gamma(X(T_m)) \leq \gamma\}, \end{aligned} \tag{11}$$

where the latter holds by (10).

Therefore, for any \mathbb{A} -stopping time $M' \geq 1$,

$$\begin{aligned} 0 = v(x, \Gamma(x)) &= \sup_{M' \geq 1} \mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} [R(X(T_k)) - \Gamma(x)] \right] \geq \mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} [R(X(T_k)) - \Gamma(x)] \right] \\ &= \mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} R(X(T_k)) \right] - \Gamma(x) \mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} \right], \end{aligned} \tag{12}$$

and thus

$$\Gamma(x) \geq \frac{\mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} R(X(T_k)) \right]}{\mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} \right]}.$$

On the other hand, by (11), the inequality (12) holds with equality when M' equals

$$\tau(x, \Gamma(x)) = \inf\{m \geq 1 : \Gamma(X(T_m)) \leq \Gamma(x)\}.$$

This implies that

$$\Gamma(x) = \sup_{M' \geq 1} \frac{\mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} R(X(T_k)) \right]}{\mathbb{E}_x \left[\sum_{k=0}^{M'-1} e^{-qT_k} \right]} = \frac{\mathbb{E}_x \left[\sum_{k=0}^{\tau(x, \Gamma(x))-1} e^{-qT_k} R(X(T_k)) \right]}{\mathbb{E}_x \left[\sum_{k=0}^{\tau(x, \Gamma(x))-1} e^{-qT_k} \right]}. \tag{13}$$

Note that both the numerator and denominator are finite by Assumption 3.1(2).

We now write (13) more explicitly. By Assumption 3.1 and because X is spatially homogeneous, $\Gamma(x)$ is strictly increasing in x and thus

$$\tau(x, \Gamma(x)) = \inf\{m \geq 1 : X(T_m) \leq x\}. \quad (14)$$

Because N is a Poisson process with arrival times $\mathcal{T} := (T_i)_{i \in \mathbb{N}}$, using

$$T_x^- := T_{\tau(x, \Gamma(x))} = \inf\{S \in \mathcal{T} : X(S) \leq x\},$$

we have

$$\begin{aligned} & \mathbb{E}_x \left[\sum_{k=0}^{\tau(x, \Gamma(x))-1} e^{-qT_k} R(X(T_k)) \right] \\ &= R(x) + \mathbb{E}_x \left[\int_{(0, T_x^-)} e^{-qs} R(X(s)) dN(s) \right] - \mathbb{E}_x \left[e^{-qT_x^-} R(X(T_x^-)) \right] \\ &= R(x) + \mathbb{E}_x \left[\int_{(0, T_x^-)} e^{-qs} R(X(s)) 1_{\{X(s) > x\}} dN(s) \right] \\ &= R(x) + \lambda \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} R(X(s)) 1_{\{X(s) > x\}} ds \right]. \end{aligned}$$

Here, the second equality holds because if $T_i < T_x^-$ then $X(T_i) > x$ for all $i \geq 1$ whereas $X(T_x^-) \leq x$ by the definition of T_x^- . The last equality holds by compensation formula.

We hence have the following.

Theorem 1. *The Gittins index Γ for Problem 1 is*

$$\Gamma(x) = \frac{R(x) + \lambda \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} R(X(s)) 1_{\{X(s) > x\}} ds \right]}{1 + \lambda \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} 1_{\{X(s) > x\}} ds \right]}, \quad x \in \mathbb{R}. \quad (15)$$

The Gittins index for Problem 2 can be written by substituting (9) in (15).

Remark 1. As $\lambda \rightarrow 0$, the Gittins index $\Gamma(x)$ converges to the instantaneous reward $R(x)$ for all $x \in \mathbb{R}$. In particular, it converges to the q -resolvent $\mathbb{E}_x [\int_0^\infty e^{-qt} r(X(t)) dt]$ for Problem 2. These results are consistent because, as λ goes to zero, the future rewards vanish.

4 Spectrally one-sided cases

In this section, we focus on the spectrally negative and positive cases and compute the Gittins index (15) in terms of the scale function. Recall that a spectrally negative

Lévy process is a Lévy process without positive jumps that does not have monotone paths a.s. The spectrally positive Lévy process is its dual process.

4.1 Review of fluctuation theory for spectrally negative Lévy processes

We first review the fluctuation theory and scale function of the spectrally negative Lévy process.

Let $(Y(t))_{t \geq 0}$ be a spectrally negative Lévy process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote its Laplace exponent by

$$\psi(\theta) := \log \mathbb{E}[e^{\theta Y(1)}], \quad \theta \geq 0, \quad (16)$$

which is known to be convex on $[0, \infty)$ and admits its right-inverse

$$\Phi(q) := \sup\{s \geq 0 : \psi(s) = q\}, \quad q \geq 0. \quad (17)$$

For $y \in \mathbb{R}$, let \mathbb{P}_y be the conditional probability under which $Y(0) = y$ and \mathbb{E}_y be its expectation operator. We omit the subscript when $y = 0$.

Fix $q \geq 0$. The q -scale function $W^{(q)}$ is a mapping from \mathbb{R} to $[0, \infty)$ that takes value zero on $(-\infty, 0)$, while on $[0, \infty)$ it is a continuous and strictly increasing function with the Laplace transform

$$\int_0^\infty e^{-\theta x} W^{(q)}(x) dx = \frac{1}{\psi(\theta) - q}, \quad \theta > \Phi(q). \quad (18)$$

Define also the *second scale function*

$$Z^{(q)}(x; \theta) := e^{\theta x} \left(1 + (q - \psi(\theta)) \int_0^x e^{-\theta z} W^{(q)}(z) dz \right), \quad x \in \mathbb{R}, \theta \geq 0.$$

In particular, for $x \in \mathbb{R}$ and $\lambda > 0$,

$$Z^{(q)}(x; \Phi(q + \lambda)) = e^{\Phi(q + \lambda)x} \left(1 - \lambda \int_0^x e^{-\Phi(q + \lambda)z} W^{(q)}(z) dz \right),$$

which can also be written by (18),

$$\begin{aligned} Z^{(q)}(x; \Phi(q + \lambda)) &= e^{\Phi(q + \lambda)x} \lambda \int_x^\infty e^{-\Phi(q + \lambda)z} W^{(q)}(z) dz \\ &= \lambda \int_0^\infty e^{-\Phi(q + \lambda)z} W^{(q)}(z + x) dz. \end{aligned} \quad (19)$$

There are a number of applications of the scale function. For example, the q -resolvent can be written

$$\mathbb{E}_y \left[\int_0^\infty e^{-qs} \mathbf{1}_A(Y(s)) ds \right] = \int_{\mathbb{R}} \mathbf{1}_A(y+u) g^{(q)}(u) du, \quad (20)$$

for any Borel set A on \mathbb{R} where

$$g^{(q)}(u) := \frac{e^{-\Phi(q)u}}{\Psi'(\Phi(q))} - W^{(q)}(-u), \quad u \in \mathbb{R}; \quad (21)$$

see [10, Theorem 2.7 (iv)].

The joint Laplace transform of the first passage time

$$\tau_0^- := \inf\{t > 0 : Y(t) < 0\} \quad (22)$$

and the overshoot $Y(\tau_0^-)$ is given by the following identity

$$\begin{aligned} H^{(q)}(y; \theta) &:= \mathbb{E}_y \left[e^{-q\tau_0^- + \theta Y(\tau_0^-)} \mathbf{1}_{\{\tau_0^- < \infty\}} \right] \\ &= Z^{(q)}(y; \theta) - \frac{\Psi(\theta) - q}{\theta - \Phi(q)} W^{(q)}(y), \quad y \in \mathbb{R}, \theta \geq 0; \end{aligned} \quad (23)$$

see, e.g., Eqn. (4.5) in [12]. In particular,

$$H^{(q)}(y; \Phi(q+\lambda)) = Z^{(q)}(y; \Phi(q+\lambda)) - \frac{\lambda}{\Phi(q+\lambda) - \Phi(q)} W^{(q)}(y), \quad y \in \mathbb{R}.$$

Similar results have been obtained for the Poisson observation case. Recall that $\mathcal{T} := (T_n)_{n \geq 1}$ is the set of jump times of an independent Poisson process with rate λ . We define for $z \in \mathbb{R}$

$$\tilde{T}_z^- := \inf\{S \in \mathcal{T} : Y(S) < z\} \quad \text{and} \quad \tilde{T}_z^+ := \inf\{S \in \mathcal{T} : Y(S) > z\}.$$

Using [19, Theorem B.1] and [13, Theorem 3.1], we have the following.

Lemma 2. *Let A be a Borel set on \mathbb{R} .*

(1) *We have*

$$\mathbb{E} \left[\int_0^{\tilde{T}_0^-} e^{-qt} \mathbf{1}_A(Y(t)) dt \right] = \frac{\Phi(q+\lambda) - \Phi(q)}{\lambda} \int_A H^{(q+\lambda)}(-u; \Phi(q)) du.$$

(2) *We have*

$$\mathbb{E} \left[\int_0^{\tilde{T}_0^+} e^{-qt} \mathbf{1}_A(-Y(t)) dt \right] = \frac{\Phi(q+\lambda) - \Phi(q)}{\lambda} \int_A H^{(q)}(u; \Phi(q+\lambda)) du.$$

4.2 Gittins index for spectrally one-sided Lévy processes

We now express the Gittins index (15) in terms of the scale function for both spectrally negative and positive cases.

Proposition 1. *Suppose X is a spectrally negative Lévy process with its Laplace exponent ψ and scale function $W^{(q)}$ as defined in Section 4.1.*

1. *The Gittins index (15) for Problem 1 can be written, for $x \in \mathbb{R}$,*

$$\Gamma(x) = \frac{\Phi(q)}{\Phi(q+\lambda)} \left(R(x) + (\Phi(q+\lambda) - \Phi(q)) \int_0^\infty R(x+y) e^{-\Phi(q)y} dy \right). \quad (24)$$

2. *The Gittins index (15) for Problem 2 can be written, for $x \in \mathbb{R}$,*

$$\Gamma(x) = \frac{\Phi(q)}{\Phi(q+\lambda)} \left[\int_{\mathbb{R}} r(x+u) g^{(q+\lambda)}(u) du + (\Phi(q+\lambda) - \Phi(q)) \int_0^\infty e^{-\Phi(q)y} \int_{\mathbb{R}} r(x+y+u) g^{(q+\lambda)}(u) du dy \right], \quad (25)$$

where $g^{(q+\lambda)}$ is defined as in (21) (with q replaced by $q+\lambda$).

Proof. (1) By Lemma 2(1) with $Y = X$ and because $H^{(q+\lambda)}(-y; \Phi(q)) = e^{-\Phi(q)y}$ for all $y \geq 0$,

$$\begin{aligned} \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} R(X(s)) 1_{\{X(s) > x\}} ds \right] &= \mathbb{E} \left[\int_0^{T_0^-} e^{-qs} R(X(s) + x) 1_{\{X(s) > 0\}} ds \right] \\ &= \frac{\Phi(q+\lambda) - \Phi(q)}{\lambda} \int_0^\infty R(x+y) e^{-\Phi(q)y} dy. \end{aligned}$$

By replacing $R(\cdot)$ with 1, we also have $\mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} 1_{\{X(s) > x\}} ds \right] = (\Phi(q+\lambda) - \Phi(q)) / (\lambda \Phi(q))$. Substituting these in (15), we have (24).

(2) By (9) and (20), we have $R(x) = \int_{\mathbb{R}} r(x+u) g^{(q+\lambda)}(u) du$, which is finite by Assumption 3.2. Substituting this in (24), we have (25).

Proposition 2. *Suppose X is a spectrally positive Lévy process whose dual process $-X$ is a spectrally negative Lévy process with its Laplace exponent ψ and scale function $W^{(q)}$.*

1. *The Gittins index for Problem 1 is given by, for $x \in \mathbb{R}$,*

$$\Gamma(x) = \frac{q\Phi(q+\lambda)}{(\lambda+q)\Phi(q)} \left(R(x) + (\Phi(q+\lambda) - \Phi(q)) \int_0^\infty R(x+y) H^{(q)}(y; \Phi(q+\lambda)) dy \right). \quad (26)$$

2. The Gittins index for Problem 2 is given by, for $x \in \mathbb{R}$,

$$\begin{aligned} \Gamma(x) &= \frac{q\Phi(q+\lambda)}{(\lambda+q)\Phi(q)} \left(\int_{\mathbb{R}} r(x-u)g^{(q+\lambda)}(u)du \right. \\ &\quad \left. + (\Phi(q+\lambda) - \Phi(q)) \int_0^\infty H^{(q)}(y; \Phi(q+\lambda)) \int_{\mathbb{R}} r(x+y-u)g^{(q+\lambda)}(u)dudy \right). \end{aligned} \quad (27)$$

Proof. (1) By Lemma 2(2) with $Y = -X$,

$$\begin{aligned} & \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} R(X(s)) 1_{\{X(s) > x\}} ds \right] \\ &= \mathbb{E} \left[\int_0^{\tilde{T}_0^+} e^{-qs} R(-Y(s) + x) 1_{\{-Y(s) > 0\}} ds \right] \\ &= \frac{\Phi(q+\lambda) - \Phi(q)}{\lambda} \int_0^\infty R(x+y) H^{(q)}(y; \Phi(q+\lambda)) dy, \end{aligned}$$

which is finite by Assumption 3.1(2). On the other hand,

$$\mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} 1_{\{X(s) > x\}} ds \right] = \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} ds \right] - \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} 1_{\{X(s) \leq x\}} ds \right]. \quad (28)$$

Here, by identity (62) in [16],

$$\begin{aligned} \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} ds \right] &= \frac{1}{q} \left(1 - \mathbb{E}_x \left[e^{-qT_x^-} \right] \right) = \frac{1}{q} \left(1 - \frac{\Phi(q+\lambda) - \Phi(q)}{\Phi(q+\lambda)} \right) \\ &= \frac{1}{q} \frac{\Phi(q)}{\Phi(q+\lambda)}, \end{aligned}$$

and, by Lemma 2(2) with $Y = -X$ and $A = (-\infty, 0]$,

$$\begin{aligned} \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} 1_{\{X(s) \leq x\}} ds \right] &= \mathbb{E} \left[\int_0^{\tilde{T}_0^+} e^{-qs} 1_{\{-Y(s) \leq 0\}} ds \right] \\ &= \int_{-\infty}^0 \frac{\Phi(q+\lambda) - \Phi(q)}{\lambda} H^{(q)}(y; \Phi(q+\lambda)) dy \\ &= \int_0^\infty \frac{\Phi(q+\lambda) - \Phi(q)}{\lambda} e^{-\Phi(q+\lambda)y} dy \\ &= \frac{\Phi(q+\lambda) - \Phi(q)}{\Phi(q+\lambda)\lambda}. \end{aligned}$$

Substituting these in (28) gives

$$\begin{aligned} 1 + \lambda \mathbb{E}_x \left[\int_0^{T_x^-} e^{-qs} \mathbf{1}_{\{X(s) > x\}} ds \right] &= 1 + \lambda \left(\frac{1}{q} \frac{\Phi(q)}{\Phi(q+\lambda)} - \frac{\Phi(q+\lambda) - \Phi(q)}{\Phi(q+\lambda)\lambda} \right) \\ &= \frac{(\lambda + q)\Phi(q)}{q\Phi(q+\lambda)}. \end{aligned}$$

Substituting these we have (26).

(2) By (9) and (20),

$$\begin{aligned} R(x) &= \mathbb{E} \left[\int_0^\infty e^{-(q+\lambda)s} r(X(s) + x) ds \right] = \mathbb{E} \left[\int_0^\infty e^{-(q+\lambda)s} r(-Y(s) + x) ds \right] \\ &= \int_{\mathbb{R}} r(x-u) g^{(q+\lambda)}(u) du, \end{aligned}$$

which is finite by Assumption 3.2(2). Substituting this in (26), we have (27).

5 Convergence of the Gittins index as $\lambda \rightarrow \infty$

In this section, we show the convergence of the Gittins indices obtained in Propositions 1 and 2 to those in the classical Lévy case [9].

5.1 Classical Lévy bandit case

We first review and obtain a characterization of the Gittins index of [9]. Let Y be a spectrally negative Lévy process and use the same notations used in Section 4.1.

For the spectrally negative case, the local time at maximum can be selected to be $L(t) = \bar{Y}(t) := \sup_{0 \leq u \leq t} Y(u)$ for $t \geq 0$. Then, the ascending ladder height process becomes $(L^{-1}(t), H(t))_{t \geq 0}$ where

$$L^{-1}(t) := \inf\{s > 0 : \bar{Y}(s) > t\} = \inf\{s > 0 : Y(s) > t\},$$

and because Y does not jump upwards,

$$H(t) := Y(L^{-1}(t)) = t,$$

(see [11, Section 6.5.2]). As in [11, (6.34)], the pair $(L^{-1}(t), H(t))_{t \geq 0}$ becomes a two-dimensional subordinator with its Laplace exponent

$$\mathbb{E}[e^{-qL^{-1}(1) - \theta H(1)} \mathbf{1}_{\{1 < L(\infty)\}}] = e^{-\kappa(q, \theta)} \quad (29)$$

where

$$\kappa(q, \theta) := \Phi(q) + \theta, \quad q, \theta \geq 0.$$

By the Wiener-Hopf factorization, with \widehat{L} and $(\widehat{L}^{-1}, \widehat{H})$ those for the dual process $-Y$ (called the descending ladder height process)

$$\mathbb{E}[e^{-q\widehat{L}^{-1}(1)-\theta\widehat{H}(1)}1_{\{1<\widehat{L}(\infty)\}}] = e^{-\widehat{\kappa}(q,\theta)} \quad (30)$$

with

$$\widehat{\kappa}(q, \theta) := \frac{q - \psi(\theta)}{\Phi(q) - \theta}, \quad q, \theta \geq 0;$$

see [11, Eqn. (6.35)].

Using these, the Laplace transform (6) that characterizes the Gittins index in the classical Lévy model can be written explicitly.

(1) Suppose X is a spectrally negative Lévy process. By (29) for $Y = X$,

$$\bar{\varphi}(q, \theta) = \kappa(q, \theta) = \Phi(q) + \theta, \quad q, \theta \geq 0,$$

and thus (6) becomes

$$\int_{(0,\infty)} e^{-\theta y} \mu(dy) = \frac{\Phi(q)}{\Phi(q) + \theta}, \quad \theta \geq 0. \quad (31)$$

(2) Suppose X is a spectrally positive Lévy process. Let ψ be the Laplace exponent of its dual (spectrally negative) Lévy process and Φ its right inverse. By (30) for $Y = -X$,

$$\bar{\varphi}(q, \theta) = \widehat{\kappa}(q, \theta) = \frac{q - \psi(\theta)}{\Phi(q) - \theta}, \quad \theta \geq 0,$$

and (6) becomes

$$\int_{(0,\infty)} e^{-\theta y} \mu(dy) = \frac{q - \theta - \Phi(q)}{\Phi(q) \psi(\theta) - q}, \quad \theta \geq 0. \quad (32)$$

5.2 Convergence of the Gittins index

Suppose X is spectrally negative. For Problem 1, the Gittins index (24) can be written

$$\Gamma(x) = \int_{(0,\infty)} R(x+y) \mu_1^\lambda(dy), \quad x \in \mathbb{R}, \quad (33)$$

with

$$\mu_1^\lambda(dy) := \frac{\Phi(q)}{\Phi(q+\lambda)} \left(\delta_0(dy) + 1_{\{y>0\}} (\Phi(q+\lambda) - \Phi(q)) e^{-\Phi(q)y} dy \right), \quad y \geq 0,$$

where δ_0 is the Dirac measure at zero.

For Problem 2, the (normalized) Gittins index (25) can be written

$$(q + \lambda)\Gamma(x) = \frac{(q + \lambda)\Phi(q)}{\Phi(q + \lambda)} \left[\int_{\mathbb{R}} r(x + u)g^{(q+\lambda)}(u)du \right. \quad (34)$$

$$\left. + (\Phi(q + \lambda) - \Phi(q)) \int_0^\infty \int_{\mathbb{R}} r(x + w)e^{-\Phi(q)y}g^{(q+\lambda)}(w - y)dwdy \right] \quad (35)$$

$$= \int_{\mathbb{R}} r(x + u)\mu_2^\lambda(du), \quad x \in \mathbb{R}, \quad (36)$$

where

$$\mu_2^\lambda(du) := \frac{(q + \lambda)\Phi(q)}{\Phi(q + \lambda)} \left[g^{(q+\lambda)}(u) + (\Phi(q + \lambda) - \Phi(q))(f^{(q)} \star g^{(q+\lambda)})(u) \right] du, \quad u \in \mathbb{R}.$$

Here,

$$(f^{(q)} \star g^{(q+\lambda)})(u) := \int_0^\infty e^{-\Phi(q)y}g^{(q+\lambda)}(u - y)dy$$

which is a convolution of $f^{(q)}(u) := e^{-\Phi(q)u}1_{\{u>0\}}$ and $g^{(q+\lambda)}$ as defined in (21).

In (36), we consider a normalized version because $R(\cdot)$ in (9) depends on λ . This normalization is appropriate in view of Remark 2(2) below.

Remark 2. 1. By setting $R \equiv 1$ in (33), we have $1 = \Gamma(x) = \mu_1^\lambda([0, \infty))$ and hence μ_1^λ is a probability measure.

2. When $r \equiv 1$, $R = (q + \lambda)^{-1}$ (see (9)) and thus $\Gamma(x) \equiv (q + \lambda)^{-1}$. Substituting these in (36), we have $\mu_2^\lambda(\mathbb{R}) = 1$, and thus μ_2^λ is a probability measure.

Proposition 3. *When X is spectrally negative, the measures μ_1^λ and μ_2^λ converge weakly to μ (defined by (31)) as $\lambda \rightarrow \infty$.*

Proof. Proof for Problem 1. For $\theta \geq 0$, because

$$\begin{aligned} & \int_{[0, \infty)} e^{-\theta y} [\delta_0(dy) + (\Phi(q + \lambda) - \Phi(q))1_{\{y>0\}}e^{-\Phi(q)y}dy] \\ &= 1 + \frac{\Phi(q + \lambda) - \Phi(q)}{\Phi(q) + \theta} = \frac{\theta + \Phi(q + \lambda)}{\Phi(q) + \theta}, \end{aligned}$$

the Laplace transform of μ_1^λ becomes

$$\int_{[0, \infty)} e^{-\theta y} \mu_1^\lambda(dy) = \frac{\Phi(q)}{\Phi(q + \lambda)} \frac{\theta + \Phi(q + \lambda)}{\Phi(q) + \theta}, \quad (37)$$

which converges to (31) as $\lambda \rightarrow \infty$. By the the continuity theorem for Laplace transforms, the weak convergence holds.

Proof for Problem 2. Because μ_2^λ has a support \mathbb{R} , we consider Fourier transforms. To this end, define the characteristic exponent of $X = Y$ by $\Psi(\theta) = -\psi(i\theta)$ such that

$$\mathbb{E}[e^{i\theta X(s)}] = e^{-\Psi(\theta)s}, \quad \theta \in \mathbb{R}, s \geq 0. \quad (38)$$

Fix $\theta \in \mathbb{R}$. Using (20) and (38), the Fourier transform of $g^{(q+\lambda)}$ becomes

$$\begin{aligned} \int_{\mathbb{R}} e^{i\theta y} g^{(q+\lambda)}(y) dy &= \mathbb{E} \left[\int_0^\infty e^{-(q+\lambda)s} e^{i\theta X(s)} ds \right] \\ &= \int_0^\infty e^{-(q+\lambda)s} e^{-\Psi(\theta)s} ds = \frac{1}{\Psi(\theta) + (q+\lambda)}. \end{aligned} \quad (39)$$

On the other hand, the Fourier transform of $f^{(q)}$ becomes $\int_{\mathbb{R}} e^{i\theta y} f^{(q)}(y) dy = \int_0^\infty e^{i\theta y} e^{-\Phi(q)y} dy = (\Phi(q) - i\theta)^{-1}$. By these and the convolution theorem,

$$\int_{\mathbb{R}} e^{i\theta y} (f^{(q)} \star g^{(q+\lambda)})(y) dy = \frac{1}{(\Psi(\theta) + (q+\lambda))(\Phi(q) - i\theta)}.$$

This together with (39) gives

$$\begin{aligned} \int_{\mathbb{R}} e^{i\theta y} \mu_2^\lambda(dy) &= \frac{\Phi(q)}{\Phi(q+\lambda)} \frac{(q+\lambda)}{\Psi(\theta) + (q+\lambda)} \left(1 + \frac{\Phi(q+\lambda) - \Phi(q)}{\Phi(q) - i\theta} \right) \\ &= \frac{\Phi(q+\lambda) - i\theta}{\Phi(q+\lambda)} \frac{(q+\lambda)}{\Psi(\theta) + (q+\lambda)} \frac{\Phi(q)}{\Phi(q) - i\theta}, \end{aligned} \quad (40)$$

which converges as $\lambda \rightarrow \infty$ to $\int_{[0, \infty)} e^{i\theta y} \mu(dy) = \Phi(q)/(\Phi(q) - i\theta)$, matching (31) (with θ replaced by $-i\theta$). By the continuity theorem, the weak convergence holds.

Remark 3. (1) For Problem 1, the Laplace transform (37) can be written

$$\int_{[0, \infty)} e^{-\theta y} \mu_1^\lambda(dy) = \frac{\varphi_1(0, q, \lambda)}{\varphi_1(\theta, q, \lambda)} \quad \text{where } \varphi_1(\theta, q, \lambda) := \frac{\Phi(q) + \theta}{\theta + \Phi(q + \lambda)}.$$

(2) For Problem 2, the Fourier transform (40) can be written

$$\int_{\mathbb{R}} e^{i\theta y} \mu_2^\lambda(dy) = \frac{\varphi_2(0, q, \lambda)}{\varphi_2(\theta, q, \lambda)} \quad \text{where } \varphi_2(\theta, q, \lambda) := \frac{(\Psi(\theta) + (q+\lambda))(\Phi(q) - i\theta)}{\Phi(q+\lambda) - i\theta}.$$

These expressions can be seen as generalizations of the classical case (6).

We now consider the spectrally positive case. For Problem 1, the Gittins index (26) can be written

$$\Gamma(x) = \int_{[0, \infty)} R(x+y) \mu_1^\lambda(dy), \quad x \in \mathbb{R},$$

with

$$\mu_1^\lambda(dy) := \frac{q\Phi(q+\lambda)}{(\lambda+q)\Phi(q)} \left(\delta_0(dy) + 1_{\{y>0\}}(\Phi(q+\lambda) - \Phi(q))H^{(q)}(y, \Phi(q+\lambda))dy \right), \quad y \geq 0.$$

For Problem 2, the Gittins index (27) can be written with $\widehat{g}(u) := g(-u)$, $u \in \mathbb{R}$,

$$\begin{aligned}
& (q + \lambda)\Gamma(x) \\
&= \frac{q\Phi(q + \lambda)}{\Phi(q)} \left[\int_{\mathbb{R}} r(x - u)g^{(q + \lambda)}(u)du \right. \\
&\quad \left. + (\Phi(q + \lambda) - \Phi(q)) \int_0^\infty \int_{\mathbb{R}} H^{(q)}(y; \Phi(q + \lambda))r(x + y - u)g^{(q + \lambda)}(u)dudy \right] \\
&= \frac{q\Phi(q + \lambda)}{\Phi(q)} \left[\int_{\mathbb{R}} r(x + u)\widehat{g}^{(q + \lambda)}(u)du \right. \\
&\quad \left. + (\Phi(q + \lambda) - \Phi(q)) \int_{\mathbb{R}} \int_0^\infty H^{(q)}(y; \Phi(q + \lambda))r(x + w)\widehat{g}^{(q + \lambda)}(w - y)dydw \right] \\
&= \int_{\mathbb{R}} r(x + u)\mu_2^\lambda(du)
\end{aligned}$$

where

$$\mu_2^\lambda(du) := \frac{q\Phi(q + \lambda)}{\Phi(q)} \left(\widehat{g}^{(q + \lambda)}(u) + (\Phi(q + \lambda) - \Phi(q))(f^{(q)} \star \widehat{g}^{(q + \lambda)})(u) \right) du, \quad u \in \mathbb{R},$$

with $f^{(q)}(u) := H^{(q)}(u, \Phi(q + \lambda))1_{\{u > 0\}}$.

Remark 4. Similar to Remark 2, both μ_1^λ and μ_2^λ are probability measures.

Proposition 4. *When X is spectrally positive, the measures μ_1^λ and μ_2^λ converge weakly to μ (defined by (32)) as $\lambda \rightarrow \infty$.*

Proof. Proof for Problem 1. In order to characterize the measure μ^λ we compute its Laplace transform. To this end, for $\theta > \Phi(q)$, using (19),

$$\begin{aligned}
& \int_0^\infty e^{-\theta y} Z^{(q)}(y; \Phi(q + \lambda)) dy \\
&= \lambda \int_0^\infty e^{-\theta y} \int_y^\infty e^{-\Phi(q + \lambda)(u - y)} W^{(q)}(u) du dy \\
&= \lambda \int_0^\infty e^{-\Phi(q + \lambda)u} W^{(q)}(u) \int_0^u e^{-(\theta - \Phi(q + \lambda))y} dy du \\
&= \frac{\lambda}{\Phi(q + \lambda) - \theta} \int_0^\infty e^{-\Phi(q + \lambda)u} W^{(q)}(u) \left(e^{(\Phi(q + \lambda) - \theta)u} - 1 \right) du \\
&= \frac{\lambda}{\Phi(q + \lambda) - \theta} \left(\int_0^\infty e^{-\theta u} W^{(q)}(u) du - \int_0^\infty e^{-\Phi(q + \lambda)u} W^{(q)}(u) du \right) \\
&= \frac{\lambda + q - \psi(\theta)}{(\Phi(q + \lambda) - \theta)(\psi(\theta) - q)}, \tag{41}
\end{aligned}$$

and hence, again using (18),

$$\begin{aligned}
& (\Phi(q+\lambda) - \Phi(q)) \int_0^\infty e^{-\theta y} H^{(q)}(y, \Phi(q+\lambda)) dy \\
&= (\Phi(q+\lambda) - \Phi(q)) \frac{\lambda + q - \psi(\theta)}{(\Phi(q+\lambda) - \theta)(\psi(\theta) - q)} - \frac{\lambda}{\psi(\theta) - q} \\
&= \frac{(q - \psi(\theta))\Phi(q+\lambda) - (\lambda + q - \psi(\theta))\Phi(q) + \lambda\theta}{(\psi(\theta) - q)(\Phi(q+\lambda) - \theta)}, \tag{42}
\end{aligned}$$

which holds for any $\theta \geq 0$ by analytic continuation and is finite in view of Lemma 2(2). Thus,

$$\begin{aligned}
& \int_{[0, \infty)} e^{-\theta y} \left[\delta_0(dy) + \mathbf{1}_{\{y>0\}} (\Phi(q+\lambda) - \Phi(q)) H^{(q)}(y, \Phi(q+\lambda)) dy \right] \\
&= 1 + \frac{(q - \psi(\theta))\Phi(q+\lambda) - (\lambda + q - \psi(\theta))\Phi(q) + \lambda\theta}{(\psi(\theta) - q)(\Phi(q+\lambda) - \theta)} \\
&= \frac{(\lambda + q - \psi(\theta))(\theta - \Phi(q))}{(\psi(\theta) - q)(\Phi(q+\lambda) - \theta)}.
\end{aligned}$$

Therefore,

$$\int_{[0, \infty)} e^{-\theta y} \mu_1^\lambda(dy) = \frac{\lambda + q - \psi(\theta)}{\lambda + q} \frac{\Phi(q+\lambda)}{\Phi(q+\lambda) - \theta} \frac{\theta - \Phi(q)}{\psi(\theta) - q} \frac{q}{\Phi(q)}, \quad \theta \geq 0, \tag{43}$$

which converges to (32) as $\lambda \rightarrow \infty$. By the continuity theorem, the weak convergence holds.

Proof for Problem 2. Again let $\Psi(\theta) = -\psi(i\theta)$ (where we recall ψ is for the dual spectrally negative Lévy process $Y = -X$) and hence $\mathbb{E}[e^{i\theta Y(s)}] = e^{-\Psi(\theta)s}$, $\theta \in \mathbb{R}, s \geq 0$.

We have

$$\begin{aligned}
\int_{\mathbb{R}} e^{i\theta y} \widehat{g}^{(q+\lambda)}(y) dy &= \mathbb{E} \left[\int_0^\infty e^{-(q+\lambda)s} e^{-i\theta Y(s)} ds \right] \\
&= \int_0^\infty e^{-(q+\lambda)s} e^{-\Psi(-\theta)s} ds = \frac{1}{\Psi(-\theta) + (q+\lambda)}. \tag{44}
\end{aligned}$$

Because (42) also holds for $\theta = 0$ and it is finite, by analytic continuation, we obtain

$$\begin{aligned}
& (\Phi(q+\lambda) - \Phi(q)) \int_0^\infty e^{i\theta y} H^{(q)}(y, \Phi(q+\lambda)) dy \\
&= \frac{(q + \Psi(-\theta))\Phi(q+\lambda) - (\lambda + q + \Psi(-\theta))\Phi(q) - i\lambda\theta}{(-\Psi(-\theta) - q)(\Phi(q+\lambda) + i\theta)}.
\end{aligned}$$

Together with (44) and the convolution theory for Fourier transforms,

$$\begin{aligned}
& \int_{\mathbb{R}} e^{i\theta y} \mu_2^\lambda(dy) \\
&= \frac{q\Phi(q+\lambda)}{\Phi(q)} \frac{1}{\Psi(-\theta) + (q+\lambda)} \left(1 + \frac{(q+\Psi(-\theta))\Phi(q+\lambda) - (\lambda+q+\Psi(-\theta))\Phi(q) - i\lambda\theta}{(-\Psi(-\theta) - q)(\Phi(q+\lambda) + i\theta)} \right) \\
&= \frac{i\theta + \Phi(q)}{\Phi(q)} \frac{q}{\Psi(-\theta) + q} \frac{\Phi(q+\lambda)}{\Phi(q+\lambda) + i\theta},
\end{aligned} \tag{45}$$

which converges as $\lambda \rightarrow \infty$ to

$$\int_{[0, \infty)} e^{i\theta y} \mu(dy) = \frac{q}{\Phi(q)} \frac{i\theta + \Phi(q)}{\Psi(-\theta) + q},$$

matching (32) (with θ replaced by $-i\theta$). Therefore the continuity theorem shows the weak convergence.

Remark 5. (1) For Problem 1, the Laplace transform (43) can be written

$$\int_{[0, \infty)} e^{-\theta y} \mu_1^\lambda(dy) = \frac{\varphi_1(0, q, \lambda)}{\varphi_1(\theta, q, \lambda)} \quad \text{where } \varphi_1(\theta, q, \lambda) := \frac{(\psi(\theta) - q)(\Phi(q+\lambda) - \theta)}{(\lambda + q - \psi(\theta))(\theta - \Phi(q))}.$$

(2) For Problem 2, the Fourier transform (45) can be written

$$\int_{\mathbb{R}} e^{i\theta y} \mu_2^\lambda(dy) = \frac{\varphi_2(0, q, \lambda)}{\varphi_2(\theta, q, \lambda)} \quad \text{where } \varphi_2(\theta, q, \lambda) := \frac{(\Psi(-\theta) + q)(\Phi(q+\lambda) + i\theta)}{i\theta + \Phi(q)}.$$

These can be seen as generalizations of (6).

Remark 6. The weak convergence as in Propositions 3 and 4 implies the convergence of the Gittins index, if $R(\cdot)$ and $r(\cdot)$ are bounded and continuous in Problems 1 and 2, respectively.

6 Concluding remarks

In this paper, we studied an extension of the MAB driven by Lévy processes [8] where decision opportunities arrive at Poisson arrival times. The Gittins index can be written analogously to the classical case (15). In particular, we obtained explicitly its form in terms of the scale function for the case of spectrally one-sided Lévy processes and showed its convergence to those in [8] as the arrival rate increases to infinity.

The studied problem is new to the best of our knowledge and there are several venues for future research.

First, in the current paper, we did not discuss the optimality of the Gittins index policy, and it is of great importance to analyze its near or exact optimality. The convergence results obtained in Section 5 suggest that it is near optimal at least

when λ is high. It is a non-trivial and crucial problem to investigate whether the optimality remains to hold generally under the Poissonian decision times.

Second, one potential application of the considered problem is to approximate the optimal solutions in the classical discrete-time model (in Section 2.1) in terms of the Gittins index policy in the considered problem by approximating deterministic decision times in terms of Poisson arrival times. This is related to Carr's randomization [4], where constant discrete decision times are approximated by Erlang random variables, or the sum of independent exponential random variables. This approximation method is conjectured to be practical and accurate in view of the existing related results in stochastic control such as [14].

Third, the obtained results are conjectured to hold for more general processes. The convergence of our Gittins index to that in (15) is easily conjectured to hold for a general Lévy process with two-sided jumps. In addition, the Gittins index may admit an analytical expression written in terms of the Wiener-Hopf factorization, even when the spectrally one-sided assumption is relaxed. It is also of great interest to generalize the results to a more general Markov process with jumps by taking advantage of the recent development of excursion theory such as [17].

References

1. Albrecher, H., Ivanovs, J., Zhou, X.: Exit identities for Lévy processes observed at Poisson arrival times. *Bernoulli*. **22**, 1364–1382 (2016)
2. Avanzi, B., Cheung, E. C., Wong, B., Woo, J.K.: On a periodic dividend barrier strategy in the dual model with continuous monitoring of solvency. *Insurance Math. Econom.* **52**, 98–113 (2013)
3. Bertoin, J.: Lévy processes. Cambridge University Press, Cambridge (1996)
4. Carr, P.: Randomization and the American put. *Rev. Financ. Stud.* **11(3)**, 597–626 (1998)
5. Czarna, I., Palmowski, Z.: Dividend problem with Parisian delay for a spectrally negative Lévy risk process. *J. Optim. Theory Appl.* **161(1)**, 239–256 (2014)
6. Gittins, J.: A dynamic allocation index for the sequential design of experiments. In *Progress in Statistics*, 241–266 (1974)
7. El Karoui, N., Karatzas, I.: Dynamic allocation problems in continuous time. *Ann. Appl. Probab.* **4(2)**, 255–286 (1994)
8. Kaspi, H., Mandelbaum, A.: Multi armed bandits driven by Lévy processes. *Ann. Appl. Probab.* **5(2)**, 541–565 (1995)
9. Kaspi, H., Mandelbaum, A.: Multi armed bandits in discrete and continuous time. *Ann. Appl. Probab.* **8(4)**, 1270–1290 (1998)
10. Kuznetsov, A., Kyprianou, A.E., Rivero, V.: The theory of scale functions for spectrally negative Lévy processes. In *Lévy Matters II*, pp. 97–186, Springer Lecture Notes in Mathematics (2013)
11. Kyprianou, A.E.: *Fluctuations of Lévy processes with applications*. Springer, Berlin (2014)
12. Kyprianou, A.E., Surya, B.A.: Principles of smooth and continuous fit in the determination of endogenous bankruptcy levels. *Finance Stoch.* **11**, 131–152 (2007)
13. Landriault, D., Li, B., Wong, J. T., Xu D.: Poissonian potential measures for Lévy risk models. *Insurance Math. Econom.* **82**, 152–166 (2018)
14. Leung, T., Yamazaki, K., Zhang, H.: An analytic recursive method for optimal multiple stopping: Canadization and phase-type fitting. *Int. J. Theor. Appl. Finance.* **18(05)**, 1550032 (2015)

15. Lkabous, M. A.: Poissonian occupation times of spectrally negative Lévy processes with applications. *Scand. Actuar. J.* **2021(10)**, 916-935 (2021)
16. Mata, D., Pérez, J.L. and Yamazaki, K.: Effects of positive jumps of assets on endogenous bankruptcy and optimal capital structure: continuous- and periodic-observation models. *SIAM J Financial Math.* **12(3)** 1112–1149 (2021)
17. Noba, K.: Generalized scale functions of standard processes with no positive jumps. *Electron. Commun. Probab.* **25**, 1-12 (2020)
18. Palmowski, Z., Pérez, J. L., Yamazaki, K.: Double continuation regions for American options under Poisson exercise opportunities. *Math. Financ.*, **31(2)**, 722-771 (2021)
19. Palmowski, Z., Pérez, J. L., Surya, B. A., Yamazaki, K.: The Leland-Toft optimal capital structure model under Poisson observations, *Finance Stoch.*, **24**, 1035–1082 (2020)
20. Pérez, J.L. and Yamazaki, K.: Mixed periodic-classical barrier strategies for Lévy risk processes. *Risks* **6(2)**, 33 (2018)
21. Pérez, J.L., Yamazaki, K., Bensoussan, A.: Optimal periodic replenishment policies for spectrally positive Lévy demand processes. *SIAM J. Control Optim.* **58(6)**, 3428–3456 (2020)
22. Puterman, M. L.: *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons (2014)
23. Snell, L.: Applications of martingale systems theorems. *Trans. Amer. Math. Soc.* **73**, 293-312 (1952)
24. Zhao, Y., Chen, P., Yang, H.: Optimal periodic dividend and capital injection problem for spectrally positive Lévy processes. *Insurance Math. Econom.*, **74**, 135-146 (2017)