The Annals of Applied Probability
1995, Vol. 5, No. 2, 541–565

LÉVY BANDITS: MULTI-ARMED BANDITS DRIVEN BY LÉVY PROCESSES

BY HAYA KASPI AND AVI MANDELBAUM

Technion–Israel Institute of Technology

Lévy bandits are multi-armed bandits driven by Lévy processes. As anticipated from existing research, Lévy bandits are optimally controlled by an index strategy: One can associate with each arm an index function of its state, and optimal strategies are those that allocate time to arms whose states have the largest index. Furthermore, the index function of an arm is calculated independently of the other arms, and the optimal reward can be expressed in terms of the indices. Somewhat less anticipated, however, is the fact that the index function of an arm, driven by a Lévy process, has a representation in terms of the decreasing ladder sets and the exit system of its Lévy driver. Moreover, the Wiener–Hopf factorization of the Lévy exponents of an arm can be used to obtain the characteristic function of some excursion law, through which the index of the arm is defined. We use this factorization to calculate explicitly index functions and optimal rewards of some interesting Lévy bandits, rediscovering along the way that local time naturally quantifies switching in continuous time.

1. Introduction. Multi-armed bandits are models of dynamic allocation of a scarce resource among competing projects. It is customary to interpret the resource as "time," which is dynamically allocated among several independent stochastic processes, each of which represents the evolution of an arm. The goal is then to find an optimal allocation strategy which maximizes, for example, cumulative reward discounted over an infinite horizon.

The present work is devoted to Lévy bandits. These are multi-armed bandits whose arms evolve as Lévy processes [7, 10]. We start in Sections 2 and 3 with a formulation of the multi-armed Lévy bandit problem and its solution (Theorem 3.1), within the framework of multiparameter processes [16, 17, 4, 5, 6]. The solution entails associating with each arm an *index function* of its state and showing the optimality of *index strategies*, namely, those that engage the arms whose index is maximal. Optimally is proved in Sections 5 and 8. Specifically, in Section 5 we establish the existence of index strategies, extending [17] and [18] to cover discontinuous processes. In Section 8 we prove optimally, relying on formula (3.7) for the value in terms of indices. Our proof differs from that in [5] and [6] in that it is based on excursion theory rather than martingale methods.

Received March 1994; revised October 1994.
AMS 1991 *subject classifications*. 60J30, 60G40, 60J55.
Key words and phrases. Lévy processes, excursions, local time, Wiener–Hopf factorization, multiparameter processes, multiarmed bandits, optional increasing path.

While optimality of index strategies is anticipated, in view of existing research on bandits [20, 3, 8], rarely is it the case that index functions can be explicitly computed, and Lévy bandits provide a host of exceptions to this. Indeed, as described in Sections 4 and proved in Section 7, the index function of a Lévy arm can be represented in terms of the decreasing ladder sets and the exit system of its driver (Section 4.2). Furthermore, the Wiener–Hopf factorization of the Lévy exponents of an arm can be used to obtain the characteristic function of some excursion law, through which its index is defined (Theorem 3.1 and Section 4.3). This enables, in Section 6, explicit computations of indices and value functions, notably for arms driven by Wiener, compound Poisson and some stable processes.

The importance of bandit problems, as well as their difficulty, stems from the fact that they embody the conflict between taking actions that yield immediate rewards, as opposed to pursuing those whose benefit is realizable only in the future (due, for example, to learning prerequisites). The scope of bandit problems is amply manifested by the three recent books on the subject: Presman and Sonin [20], Berry and Fristedt [3] and Gittins [8]. Quoting freely from [20]: "while these three books have equal rights to use the term bandit, the intersection of the content of any pair out of the three is either empty or rather small." Our models are closest to [11], with some overlap with [3] (especially Chapter 8). Our results can be viewed as a rare contribution of Markovian excursion theory to applied probability. There is further potential for practical applications, in view of the fundamental significance of Lévy processes (they are "the" continuous-time random walks), combined with our explicit expressions for Lévy indices.

2. Problem formulation. A multi-armed Lévy bandit consists of d individual arms, indexed by $k = 1, \ldots, d$. The primitives for arm k are given by X^k, $(\Omega^k, \mathscr{F}^k, P^k)$ and r^k, described as follows. The *state* process $X^k = \{X_t^k, t \geq 0\}$ is a real-valued stochastic process on the filtered space $(\Omega^k, \mathscr{F}^k)$; the *information* process $\mathscr{F}^k = \{\mathscr{F}_t^k, t \geq 0\}$ is a filtration in Ω^k which satisfies the usual conditions and to which X^k is adapted; $P^k = \{P_{x^k}^k, x^k \in \mathbb{R}^1\}$ is a family of probability measures on $(\Omega^k, \mathscr{F}_\infty^k)$, such that under $P_{x^k}^k$, the process X^k is a Lévy process starting from $X_0^k = x^k$; finally, the reward function $r^k(x^k)$ is real-valued, nondecreasing, continuous and for which

$$P_{x^k}^k \int_0^\infty e^{-\beta t} |r^k(X_t^k)| \, dt < \infty \qquad \forall\, x_k \in \mathbb{R}^1, \tag{2.1}$$

where $\beta > 0$ is a given discount rate. (Here and in the sequel, Pf denotes the integral of a measurable function f with respect to the measure P.)

The evolution of the multi-armed bandit is described in terms of the multiparameter process

$$X := \left\{ X_s = \left(X_{s_1}^1, \ldots, X_{s_d}^d \right), s = (s_1, \ldots, s_d) \geq 0 \right\},$$

adapted to the multiparameter filtration

$$\mathscr{F} := \{\mathscr{F}_s = \mathscr{F}_{s_1}^1 \vee \cdots \vee \mathscr{F}_{s_d}^d,\ s \geq 0\}.$$

Specifically, $X_s = (X_{s_1}^1, \ldots, X_{s_d}^d)$ is the state of the bandit after s_k units of time have been allocated to each arm k, $1 \leq k \leq d$, at which instant the information accumulated is given by $\mathscr{F}_s = \mathscr{F}_{s_1}^1 \vee \cdots \vee \mathscr{F}_{s_d}^d$. Formally, the multiparameter process X is constructed on the filtered product space $(\Omega, \mathscr{F}) = \otimes_{k=1}^d (\Omega^k, \mathscr{F}_\infty^k)$, equipped with the family of product probability measures $\{P_x, x \in \mathbb{R}^d\}$, given by $P_x = \otimes_{k=1}^d P_{x^k}^k$, $x = (x^1, \ldots, x^d)$. Thus, under each P_x, $x \in \mathbb{R}^d$, the stochastic processes X^k, $k = 1, \ldots, d$, are mutually independent and $X_0 = x$.

The evolution of the bandit is prescribed by an *allocation strategy* T, which is a d-dimensional stochastic process $T = \{T(t),\ t \geq 0\}$, where

$$T(t) = (T_1(t), \ldots, T_d(t)), \qquad t \geq 0,$$

has the following properties:

1. $T(t)$ is nondecreasing in $t \geq 0$, with $T(0) = 0$.
2. $T_1(t) + \cdots + T_d(t) = t$, for all $t \geq 0$.
3. $\{T(t) \leq s\} \in \mathscr{F}_s$, for all $t \geq 0$ and $s \geq 0$.

The random variable $T_k(t)$ stands for the total amount of time allocated to arm k over the interval $[0, t]$. Properties 1 and 2 are thus self-explanatory. Property 3 is a mathematical articulation of the nonanticipative nature of T: For all k, the event "no more that s_k units of time have been allocated to arm k" does not depend on information beyond $\mathscr{F}_{s_k}^k$. Formally,

$$\{T_1(t) \leq s_1, \ldots, T_d(t) \leq s_d\} \in \mathscr{F}_{s_1}^1 \vee \cdots \vee \mathscr{F}_{s_d}^d \qquad \forall\, t \geq 0, \forall\, (s_1, \ldots, s_d) \geq 0.$$

In the theory of multiparameter processes, $T(t)$ is called a stopping point in $\mathbb{R}^d$. An allocation strategy is called an *optional increasing path* [24]; being a nondecreasing family of stopping points, it is also referred to as a multiparameter random time change.

Under a strategy T, the state of the bandit at time $t \geq 0$ is given by the random vector

$$X_T(t) = \left(X_{T_1(t)}^1 \ldots, X_{T_d(t)}^d\right),$$

and the information available then is the σ-field

$$\mathscr{F}_T(t) = \{B \in \mathscr{F}\colon B \cap \{T(t) \leq s\} \in \mathscr{F}_s, \forall\, s \geq 0\}.$$

[The common notation for $\mathscr{F}_T(t)$ is $\mathscr{F}_{T(t)}$, which is the σ-field of events prior to the stopping point $T(t)$.] Associated with T is its present value $R(T)$ of cumulative discounted future rewards. This is the random variable

$$R(T) = \int_0^\infty e^{-\beta t} r[X_T(t)] \cdot dT(t),$$

where $r[x] = (r^1(x^1), \ldots, r_d(x^d))$ and $r[X_T(t)] \cdot dT(t)$ is an abbreviation for $\sum_{k=1}^d r^k[X_{T_k(t)}^k]\, dT_k(t)$. The *multi-armed bandit problem* is to identify optimal

strategies that maximize the expected reward. Formally, one seeks T^* that attains

$$v(x) = \sup_T P_x R(T), \qquad x \in \mathbb{R}^d,$$

over all strategies T. The function v is called the *value* function of the bandit. It must be finite in view of (2.1).

3. Solution. Optimal strategies will be described in terms of dynamic allocation indices: with each arm k one associates an index function of its state, and optimal strategies always engage arms whose state has the highest index. We now proceed with a formal description, starting with the ingredients that make up the index functions.

Fix an arm $k \in \{1, \dots, d\}$. Introduce the lower envelope $\underline{X}^k$ of its state X^k by

$$\underline{X}_t^k = \inf_{0 \le u \le t} X_u^k, \qquad t \ge 0.$$

The *excursion* process of X^k from $\underline{X}^k$ is $\xi^k = X^k - \underline{X}^k$. Excursion times are the elements in the *complement* of the weak-descending ladder set

$$M^k = \text{Closure}\{t \ge 0 \colon \xi_t^k = 0\}. \tag{3.1}$$

They constitute an open set, which is a countable union of disjoint open *excursion intervals*.

The set M^k is regenerative [15]. As such, it is either a.s. perfect or a.s. discrete and it admits a local time L^k. (A perfect set is closed and dense in itself; a discrete set consists of isolated points. The construction of local times is given, for example, in [7] or [13]. Here we just note that a local time at a point of a Markov process is unique up to a positive constant, its inverse is a subordinator and its sample paths are continuous for perfect sets and discontinuous for discrete ones.) Let τ^k denote the right-continuous inverse of L^k. Then the process $U^k = (\tau^k, -X_{\tau^k}^k)$ is a two-dimensional subordinator. Its Laplace exponent φ^k is defined through

$$\exp\bigl[t\varphi^k(\beta, \gamma)\bigr] = P_0^k \exp\bigl[-\beta\tau_t^k + \gamma X_{\tau_t^k}^k\bigr], \qquad \beta, \gamma > 0,$$

and it is given by

$$\varphi^k(\beta, \gamma) \propto \exp\left[\int_{0+}^{\infty} dt \frac{1}{t} \int_{-\infty}^{0} P_0^k\{X_t^k \in dy\}(e^{-t} - e^{-\beta t + \gamma y})\right].$$

Here $\propto$ indicates that φ^k is determined up to a positive constant, as a consequence of the same fact for L^k and its inverse τ^k.

The *index function* Γ^k of arm k is given by

$$\Gamma^k(x^k) = \int_0^{\infty} r^k(x^k + y)\mu^k(dy), \tag{3.2}$$

where μ^k is the probability measure determined by the characteristic function

$$\int_0^\infty e^{i\theta y}\mu^k(dy) = \frac{\varphi^k(\beta, i\theta)\beta}{\varphi^k(\beta,0)\big[\beta + \psi^k(\theta)\big]}, \qquad \theta \in \mathbb{R}^1, \tag{3.3}$$

and ψ^k is the Lévy exponent of X^k, namely, $\exp[-t\psi^k(\theta)] = P_0^k \exp[i\theta X_t^k]$.

Several alternative expressions for the index, some more familiar and others computationally more convenient, will be provided in the next section. The concreteness of (3.2) and (3.3) stems from the Lévy nature of the arms.

A strategy $T^* = (T_1^*, \dots, T_d^*)$ is an *index strategy* if each $T_k^* = \{T_k^*(t), t \geq 0\}$ right-increases at time $t \geq 0$ only when

$$\Gamma^k\big[X_{T^*}^k(t)\big] = \bigvee_{j=1}^d \Gamma^j\big[X_{T^*}^j(t)\big], \tag{3.4}$$

and M^k includes all the times that are left-increases of T^k but not right-increases. This definition mathematically articulates two properties: first, T^* follows the leader among the indices and second, it does not switch arms within interiors of excursion intervals.

THEOREM 3.1 (Optimality). *Consider a multi-armed Lévy bandit as described above. Then index strategies exist, they are all optimal and their common value function is given by*

$$v(x) = P_x\int_0^\infty e^{-\beta t} \bigvee_{k=1}^d \Gamma^k\big[\underline{X}_{T^*}^k(t)\big]\, dt, \qquad x \in \mathbb{R}^d, \tag{3.5}$$

where T^* is any index strategy.

The theorem is proved in Sections 5 and 8. Specifically, in Section 5.1 we establish the existence of an index strategy, by constructing a strategy T^* that *follows the leader* among the *index processes* $\Gamma^k = \Gamma^k[X_{T^*}^k]$, $k = 1, \dots, d$, while assigning priority to Γ^j over Γ^k for $j < k$. Formally, each T_k^* increases at $t \geq 0$ only when (3.4) prevails, and whenever T_l^* increases at t and $\Gamma_{T^*}^k(t) = \bigvee_{j=1}^d \Gamma_{T^*}^j(t)$ for some $k < l$, then $\Gamma_{T^*}^k$ decreases at t. Optimality is proved in Section 8, in two steps. First, it is shown that

$$P_x R(T) \leq P_x\int_0^\infty e^{-\beta t} \bigvee_{k=1}^d \Gamma^k\big[\underline{X}_{T^*}^k(t)\big]\, dt, \qquad x \in \mathbb{R}^d, \tag{3.6}$$

for any strategy T. Then, optimality of T^* is an immediate consequence of

$$P_x R(T^*) = P_x\int_0^\infty e^{-\beta t} \bigvee_{k=1}^d \Gamma^k\big[\underline{X}_{T^*}^k(t)\big]\, dt, \qquad x \in \mathbb{R}^d, \tag{3.7}$$

which is verified for any index strategy T^*.

4. Index representations. The index function in (3.2) admits several representations, five of which will now be presented. Along the way emerges an intrinsic relation with excursion theory and, consequently, with the Wiener–Hopf factorization. Note that, in view of (3.2), the index of an arm is an operator on its reward function that inherits properties such as monotonicity and continuity.

4.1. *Gittins–Jones.* The first two representations arise in a general Markovian setting, in particular Lévy processes:

$$\begin{aligned}\Gamma^k(x^k) &= \sup_{\tau>0} \frac{P^k_{x^k}\int_0^\tau e^{-\beta u} r^k\big[X^k_u\big]\,du}{P^k_{x^k}\int_0^\tau e^{-\beta u}\,du} \\ &= \lim_{\varepsilon\downarrow 0} \frac{P^k_{x^k}\int_0^{\tau_\varepsilon} e^{-\beta u} r^k\big[X^k_u\big]\,du}{P^k_{x^k}\int_0^{\tau_\varepsilon} e^{-\beta u}\,du}, \qquad x^k \in \mathbb{R}^1,\end{aligned} \tag{4.1}$$

where the supremum is over all stopping times τ with respect to $\mathscr{F}^k$ and

$$\tau_\varepsilon = \inf\big\{t > 0\colon \Gamma^k\big[X^k(t)\big] < \Gamma^k\big[X^k(0)\big] - \varepsilon\big\}.$$

The sup representation is the familiar Gittins–Jones dynamic allocation index (Gittins and Jones [9]; Gittins [8]). It follows from the fifth representation in (4.7) below and it reduces to the lim representation when the reward function r^k is monotone. (This last fact is established in [17] for continuous processes, but the proof applies to our case as well.) Neither representation is actually used here.

4.2. *Excursions.* The third representation (4.3) is based on exit systems [14] associated with the excursions ξ^k. This representation is the one that is applied in Section 8 to prove optimality of index strategies.

Let

$$D^k_t = \inf\{u > t\colon u \in M^k\}, \qquad R^k_t = D^k_t - t, \qquad t \geq 0,$$

and

$$G^k = \big\{t \geq 0\colon R^k_{t-} = 0,\ R^k_t > 0\big\}, \qquad R^k = R^k_0.$$

Elements of G^k are beginnings (left endpoints) of excursion intervals, and R^k is the first hitting time of M^k. Theorem 3 in [12], adapted to our context, guarantees the existence of a σ-finite excursion measure $\hat{P}^k$ on $\mathscr{F}^k_\infty$ and a scalar $l^k \geq 0$, such that $\hat{P}^k(R^k = 0) = 0$, and for every stochastic process Z, bounded and predictable with respect to $\{\mathscr{F}^k_{D^k_t},\ t \geq 0\}$, and every function f, bounded and measurable, the following two relations prevail:

$$P^k_{x^k}\sum_{u\in G^k} Z_u f\big(X^k_{(u+\cdot)\wedge D^k_u}\big) = P^k_{x^k}\int_0^\infty Z_t \hat{P}^k_{X^k_t} f\big(\xi^k_{\cdot\wedge R^k}\big)\,dL^k_t, \qquad x^k \in \mathbb{R}^1, \tag{4.2}$$

and

$$\int_0^t 1_{M^k}(u)\,du = l^k \cdot L_t^k, \qquad t \geq 0,$$

where

$$\hat{P}_{x^k}^k f\big(\xi_{\cdot\wedge r^k}^k\big) = \hat{P}^k f\big(x^k + \xi_{\cdot\wedge R^k}^k\big).$$

(This last relation follows from a similar one between $P_{x^k}^k$ and P_0^k that expresses the spacial homogeneity of a Lévy process.) The third representation of the index is

$$\Gamma^k(x^k) = \frac{l^k r^k(x^k) + \hat{P}^k \int_0^{R^k} e^{-\beta t} r^k\big(x^k + \xi_t^k\big)\,dt}{l^k + \hat{P}^k \int_0^{R^k} e^{-\beta t}\,dt}, \qquad x^k \in \mathbb{R}^1. \tag{4.3}$$

In Section 7, it is shown to coincide with the familiar (4.1), by checking its equality with (4.7) below. It also implies (3.2) and (3.3) in a way that typifies our use of excursion theory and hence will now be presented.

As a start, (4.3) gives rise to (3.2), with μ^k that is characterized by

$$\mu^k(f) = \frac{l^k f(0) + \hat{P}^k \int_0^{R^k} e^{-\beta t} f\big(\xi_t^k\big)\,dt}{l^k + \hat{P}^k \int_0^{R^k} e^{-\beta t}\,dt},$$

f bounded and measurable. To deduce (3.3), recall that ψ^k denotes the Lévy exponent of X^k. Then calculate

$$P_0^k \int_0^\infty \exp(-\beta t)\exp\big(i\theta X_t^k\big)\,dt = \frac{1}{\beta + \psi^k(\theta)}.$$

Next, observe that also

$$\begin{aligned} &P_0^k \int_0^\infty \exp(-\beta t)\exp\big(i\theta X_t^k\big)\,dt \\ &\quad = P_0^k \sum_{u \in G^k} \exp(-\beta u) \int_u^{D_u^k} \exp(-\beta(t-u))\exp\big(i\theta X_t^k\big)\,dt \\ &\qquad + P_0^k \int_0^\infty \exp(-\beta t) 1_{M^k}(t) \exp\big(i\theta X_t^k\big)\,dt. \end{aligned}$$

Thus, applying (4.2) with $Z_u = e^{-\beta u}$ and

$$f\big(X_{(u+\cdot)\wedge D_u^k}^k\big) = \int_u^{D_u^k} \exp(-\beta(t-u))\exp\big(i\theta X_t^k\big)\,dt$$

yields

$$
\begin{aligned}
&P_0^k \int_0^\infty \exp(-\beta t)\exp\bigl(i\theta X_t^k\bigr)\,dt \\
&\quad = P_0^k \int_0^\infty \exp(-\beta t)\exp\bigl(i\theta X_t^k\bigr)\Bigl(l^k + \hat{P}^k \int_0^{R^k} \exp(-\beta u)\exp\bigl(i\theta \xi_u^k\bigr)\,du\Bigr)\,dL_t^k \\
&\quad = P_0^k \int_0^\infty \exp\bigl(-\beta \tau_t^k\bigr)\exp\bigl(i\theta X_{\tau_t}^k\bigr)\Bigl(l^k + \hat{P}^k \int_0^{R^k} \exp(-\beta u)\exp\bigl(i\theta \xi_u^k\bigr)\,du\Bigr)\,dt \\
&\quad = \frac{l^k + \hat{P}^k \int_0^{R^k} \exp(-\beta u)\exp\bigl(i\theta \xi_u^k\bigr)\,du}{\varphi^k(\beta, i\theta)},
\end{aligned}
$$

where the second equality is a consequence of a time change and the last equality is due to $P_0^k \exp(-\beta\tau_t^k + i\theta X_{\tau_t}^k) = \exp(-t\varphi^k(\beta, i\theta))$. Therefore,

$$
l^k + \hat{P}^k \int_0^{R^k} \exp(-\beta u)\exp\bigl(i\theta \xi_u^k\bigr)\,du = \frac{\varphi^k(\beta, i\theta)}{\beta + \psi^k(\theta)};
$$

revealing the characteristic measure of μ^k to be as in (3.3).

4.3. *Wiener–Hopf.* The fourth representation amounts to a simplification of (3.3) to (4.5) below, while leaving (3.2) intact. This representation is the one most convenient for computations.

Let $\overline{X}^k$ stand for the upper envelope of X^k, in terms of which one defines the strict-ascending ladder set

$$
\overline{M}^k = \text{Closure}\bigl\{t > 0\colon \overline{X}_t^k = X_t^k > X_u^k,\ 0 \le u < t\bigr\}. \tag{4.4}
$$

The set $\overline{M}^k$, similarly to M^k, is regenerative and it admits a local time $\overline{L}^k$. Let $\overline{\tau}^k$ denote the right-continuous inverse of $\overline{L}^k$. Then $\overline{U}^k = (\overline{\tau}^k, X^k(\overline{\tau}^k))$ is a two-dimensional subordinator, with Laplace exponent $\overline{\varphi}^k$ given by

$$
\overline{\varphi}^k(\beta, \gamma) \propto \exp\left[\int_{0+}^\infty dt \frac{1}{t} \int_{0+}^\infty P_0^k\bigl\{X_t^k \in dy\bigr\}\bigl(e^{-t} - e^{-\beta t - \gamma y}\bigr)\right].
$$

The index function Γ^k is again given by (3.2), but μ^k is now characterized by its Laplace transform

$$
\int_0^\infty e^{-\gamma y}\mu^k(dy) = \frac{\overline{\varphi}^k(\beta, 0)}{\overline{\varphi}^k(\beta, \gamma)}, \qquad \gamma > 0. \tag{4.5}
$$

The derivation of this simplified representation starts with the Wiener–Hopf factorization of X^k, which asserts that

$$
\overline{\varphi}^k(\beta, -i\theta)\varphi^k(\beta, i\theta) \propto \bigl[\beta + \psi^k(\theta)\bigr]. \tag{4.6}
$$

Now substituting (4.6) into (3.3) yields (4.5).

4.4. *Whittle–Weber.* The fifth and our last representation (4.7) of the index is Weber's modification [25] of Whittle [26]. Define

$$v^k(x^k, \gamma) = \sup_{\tau > 0} P^k_{x^k} \int_0^\tau e^{-\beta u} \big(r^k[X^k_u] - \gamma\big)\, du, \qquad \gamma \in \mathbb{R}^1,$$

where the supremum is over all stopping times τ with respect to $\mathscr{F}^k$. Then

$$\Gamma^k(x^k) = \inf\{\gamma > 0 \colon v^k(x^k, \gamma) = 0\}, \qquad x^k \in \mathbb{R}^1. \tag{4.7}$$

This representation is not used in our proofs. Nevertheless, we show in Section 7 that it coincides with (4.3) in order to place our results in the context of existing literature. The verification is carried out through discrete approximations of X^k by random walks. One first verifies that the function $x^k \to v^k(x^k, \gamma)$ is increasing lower semicontinuous. It then follows that $v^k(x^k, \gamma)$ is attained by a hitting time of a set of the form $(-\infty, b)$, $b \in \mathbb{R}^1$, which enables our use of excursion theory. (This special structure is lost once the reward functions need not be monotone. Difficulties then arise which have been resolved only in special cases, e.g., Brownian bandits with bitonic rewards [19].)

5. Index strategies. Let $\Gamma^1, \dots, \Gamma^d$ be stochastic processes on a common probability space, whose sample paths are right continuous with left limits. In Section 5.1, we prove the existence of a strategy T^* that *follows the leader* among them, and which assigns priority to Γ^j over Γ^k whenever $j < k$ (precise definitions will be provided as the need arises). Applying this existence result to $\Gamma^k = \Gamma^k[X^k]$ yields an index strategy.

The evolution of the *switched* process $\Gamma_{T^*} = (\Gamma^1_{T^*_1}, \dots, \Gamma^d_{T^*_d})$ has an illuminating sample-path decomposition, which is depicted in Section 5.2. The general setup is then specialized, in 5.3, to symmetric Lévy bandits. There, from the sample-path decomposition naturally emerges the role of local time in quantifying continuous-time switching.

5.1. *Following the leader. Two processes.* The existence of a strategy that follows the leader between two processes, say Γ^1 and Γ^2, is established as in [18]. Specifically, let $\mathscr{F}^k$ denote here the complete right-continuous filtration generated by Γ^k. Assume without loss that $\Gamma^1_0 = \Gamma^2_0$ and introduce

$$D = \big\{(s_1, s_2) \colon \underline{\Gamma}^1_{s_1} \le \underline{\Gamma}^2_{s_2}\big\}.$$

($\underline{\Gamma}^k$ is the lower envelope of Γ^k.) Then the closure $\overline{D}$ of D has the following three properties:

1. $\{(s_1, 0) \colon s_1 \ge 0\} \subset \overline{D}$.
2. $(s_1, s_2) \in \overline{D} \Rightarrow \{(u_1, u_2) \colon u_1 \ge s_1,\ 0 \le u_2 \le s_2\} \subset \overline{D}$.
3. $\{s \in \overline{D}\} \in \mathscr{F}^1_{s_1} \vee \mathscr{F}^2_{s_2} = \mathscr{F}_s$, $s = (s_1, s_2)$.

By Theorem 2.7 of Walsh [24], the northwest boundary of $\overline{D}$ can be parametrized as a strategy $\tau = (\tau_1, \tau_2)$, adapted to $\{\mathscr{F}_s,\ s \geq 0\}$, such that τ_2 increases at $t \geq 0$ only when

$$\underline{\Gamma}^2[\tau_2(t)] \geq \underline{\Gamma}^1[\tau_1(t)] > \underline{\Gamma}^1[\tau_1(u)] \qquad \forall\ u > t.$$

We have thus constructed a strategy τ that follows the leader between $\underline{\Gamma}^1$ and $\underline{\Gamma}^2$, and which assigns priority to $\underline{\Gamma}^1$ over $\underline{\Gamma}^2$. As in Theorem 12 of [17], τ also follows the leader between Γ^1 and Γ^2.

Induction. Suppose that there exists a strategy $\tau = (\tau_1, \ldots, \tau_{d-1})$ which follows the leader among $\Gamma^1, \ldots, \Gamma^{d-1}$, while assigning priorities to Γ^j over Γ^k for $j < k$. Define

$$\tilde{\Gamma}^1 = \bigvee_{j=1}^{d-1} \Gamma_\tau^j \in \mathscr{F}_\tau = \tilde{\mathscr{F}}^1,$$

$$\tilde{\Gamma}^2 = \Gamma^d \in \mathscr{F}^d = \tilde{\mathscr{F}}^2.$$

(Note that the sample paths of $\tilde{\Gamma}^1$ are created by concatenating the active segments of $\Gamma^k_{\tau_k}$, $k = 1, \ldots, d-1$.) From the first induction step, there exists a strategy (η_1, η_2) that follows the leader between $(\tilde{\Gamma}^i, \tilde{\mathscr{F}}^i)$, $i = 1, 2$. The sought-after strategy $T^* = (T_1^*, \ldots, T_d^*)$ is now given by

$$T_k^* = \tau_k(\eta_1), \qquad k = 1, \ldots, d-1; \qquad T_d^* = \eta_2.$$

Indeed, T^* enjoys properties 1 and 2 of a strategy. Next, each T_k^* increases at $t \geq 0$ only when

$$\Gamma^k_{T^*}(t) = \bigvee_{j=1}^{d} \Gamma^j_{T^*}(t),$$

and whenever T_l^* increases at t and $\Gamma^k_{T^*}(t) = \vee_{j=1}^{d} \Gamma^j_{T^*}(t)$ for some $k < l$, then $\Gamma^k_{T^*}$ decreases at t. Finally, the verification of property 3 starts with the observation

$$\begin{aligned} &\{T^*(t) < s\} \\ &\quad = \bigcup_u \{\tau_k(u) < s_k,\ k = 1, \ldots, d-1;\ \eta_1(t) < u,\ \eta_2(t) < s_d\} \end{aligned} \tag{5.1}$$

for any $s = (s_1, \ldots, s_d)$ and u running over the rationals in $[0, s_1 + \cdots + s_{d-1})$. The collection

$$\mathscr{G} = \left\{B\colon B \cap \{\tau(u) < (s_1, \ldots, s_{d-1})\} \in \mathscr{F}^1_{s_1} \vee \cdots \vee \mathscr{F}^d_{s_s}\right\}$$

is a σ-field that contains both $\mathscr{F}_\tau(u)$ and $\mathscr{F}^d_{s_d}$, hence also $\mathscr{F}_\tau(u) \vee \mathscr{F}^d_{s_d}$.

The fact that (η_1, η_2) is a strategy with respect to $\mathscr{F}_\tau \vee \mathscr{F}^d$ implies that the particular set $B = \{\eta_1(t) < u,\ \eta_2(t) < s_d\} \in \mathscr{F}_\tau(u) \vee \mathscr{F}^d_{s_d} \subset \mathscr{G}$. Hence each element of the union in (5.1) belongs to $\mathscr{F}_s$, and we are done.

5.2. *Sample-path decomposition.* Let T^* follow the leader among $\Gamma^1, \ldots, \Gamma^d$. The sample paths of the switched process Γ_{T^*} have a simple

description in terms of the following d processes, all of which are d-dimensional:

$$\mathscr{E}^1 = (\Gamma^1, \underline{\Gamma}^1, \ldots, \underline{\Gamma}^1), \ldots, \mathscr{E}^d = (\underline{\Gamma}^d, \ldots, \underline{\Gamma}^d, \Gamma^d).$$

Each $\mathscr{E}^k$ moves down along the diagonal when $\Gamma^k = \underline{\Gamma}^k$, executing positive excursions in parallel to the k-axes whenever $\Gamma^k > \underline{\Gamma}^k$. Its sample paths chart the motion of Γ_{T^*} when Γ^k is active (T_k^* increases). A graphical representation (see Figure 1) is illuminating for $d = 2$.

The motion of Γ_{T^*} has two components: an excursion part, in parallel to one of the axes, when only the single leader is active, and the motion down the diagonal when "several processes share leadership." This last description is only intuitive. For example, when Γ^1 and Γ^2 are two i.i.d. Brownian motions, the times of mutual leadership coincide with the zero set of another Brownian motion; hence its Lebesgue measure vanishes. We expand this example in the next subsection.

5.3. *Symmetric bandits.* A *symmetric bandit* is one whose arms are i.i.d., with reward functions all of which coincide. The index functions of the arms are also equal to each other; hence an index strategy follows the leader among the state processes. Identifying an optimal strategy is, therefore, trivial. Nevertheless, the processes that it gives rise to are quite subtle, as will be demonstrated momentarily.

It was noted in [17] that if $X^1, \ldots, X^d$ are independent standard Brownian motions, then $\sum_{k=1}^d X_T^k(t)$ is as well, for *any* strategy T. In fact, the following statement holds.

THEOREM 5.1. *Let $X^1, \ldots, X^d$ be i.i.d. Lévy processes, with $X_0^k \equiv 0$. Then for any allocation strategy $T = (T_1, \ldots, T_d)$, the process $S = \{S_t,\ t \geq 0\}$, given by*

$$S_t = \sum_{k=1}^d X_{T_k}^k(t),$$

is a Lévy process with the same distribution as X^k, $k = 1, \ldots, d$.

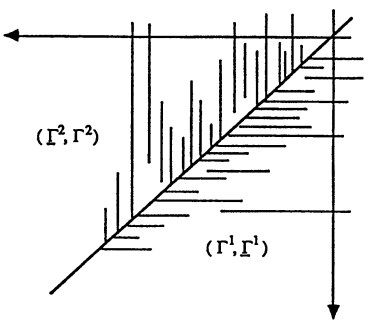

FIG. 1. *Sample path of Γ_{T^*}.*

REMARK. With a self-explanatory deviation from our notation, the theorem asserts the following additivity property: $\sum_{k=1}^{d} X^k[T_k] = X^1[\sum_{k=1}^{d} T_k]$, with equality in distribution.

PROOF. We are going to use the following characterization: X^k is a Lévy process with a Lévy exponent $\psi(\theta)$ if and only if the process $M_t^k = \exp[i\theta X_t^k + t\psi(\theta)]$, $t \geq 0$, is a complex-valued martingale for all real θ ([10], Corollary 4.13 on page 107, combined with Theorem 2.42 on page 86).

It suffices, therefore, to verify that $\exp[i\theta S_t + t\psi(\theta)]$, $t \geq 0$, is also a martingale for all real θ. To this end, we make the following two observations:

$$\left\{M_{s_1}^1 \times \cdots \times M_{s_d}^d\right\} \text{ is a multiparameter martingale}$$

by the independence of $M^1, \ldots, M^d$;

$$\left\{M_{T_1(t)}^1 \times \cdots \times M_{T_d(t)}^d\right\} \text{ is a martingale for any strategy } (T_1, \ldots, T_d),$$

as a multiparameter time change of a bounded multiparameter martingale [24]. However, this last product is precisely $\exp[i\theta S_t + t\psi(\theta)]$, since $\sum_{k=1}^{d} T_k(t) = t$, and we are done. □

THEOREM 5.2 (Path decomposition). *Consider a symmetric bandit whose arms $X^1, \ldots, X^d$ perform no downward jumps. Then their lower envelopes are continuous and a strategy T^* that follows the leader among them must satisfy*

$$\underline{X}_{T_1^*}^1 = \underline{X}_{T_2^*}^2 = \cdots = \underline{X}_{T_d^*}^d, \tag{5.2}$$

at all times. Furthermore, the sample paths of the switched process X_{T^*} have the representation

$$X_{T^*} = \xi_* - \left(\frac{1}{d}L\right) \cdot 1, \tag{5.3}$$

where ξ_* is the excursion process of X_{T^*} away from the diagonal in $\mathbb{R}^d$, whose coordinates are $\xi_*^k = \xi_{T_k^*}^k = X_{T_k^*}^k - \underline{X}_{T_k^*}^k$, $k = 1, \ldots, d$, and L is the local time of the Markov process ξ_* at the origin, where in fact $L = -\underline{S}$.

REMARK. The proof reveals that $L = -\underline{S}$ is also a local time at 0 of the Markov process $S - \underline{S}$.

PROOF. No downward jumps implies continuity of the lower envelopes which, in turn, gives rise to (5.2), as in [17]. One deduces that

$$\underline{S} = \sum_{k=1}^{d} \underline{X}_{T_k^*}^k = d \cdot \underline{X}_{T_1^*}^1, \tag{5.4}$$

where the first equality manifests the (peculiar) fact that summation and taking lower envelopes commute in our case. We have thus obtained (5.3) with $L = -\underline{S}$, and we proceed with identifying the latter as the local time of ξ_*.

By Theorem 5.1, S is a Lévy process with no downward jumps. Hence, it is a semimartingale with the decomposition

$$S_t = ct + \sigma B_t + J_t, \qquad t \geq 0, \tag{5.5}$$

where $c \in \mathbb{R}^1$, B is a standard Brownian motion, $\sigma \geq 0$ and J is a nondecreasing jump process, independent of B [7]. Introduce the semimartingale $V = S - \underline{S}$, note that $-\underline{S}$ increases only on the set $\mathscr{Z} = \{t\colon V_t = 0\}$, and apply to V Tanaka's formula ((1) on page 20 of [1]) to get

$$\begin{aligned} S_t - \underline{S}_t = {} & \int_0^t 1_{\{S_{u-} - \underline{S}_{u-} > 0\}}\, dS_u \\ & + \sum_{0 \leq u \leq t} 1_{\{S_{u-} - \underline{S}_{u-} = 0\}}\, \Delta S_u + \tfrac{1}{2} L_t^0, \qquad t \geq 0. \end{aligned} \tag{5.6}$$

Here ΔS_u is the jump of S at u and L^0 is the Tanaka local time at 0 of $S - \underline{S} = V$. Now observe that the continuous martingale part of $S - \underline{S}$ is σB. Hence, by Corollary 2 on page 32 of [1],

$$L_t^0 = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t 1_{\{S_u - \underline{S}_u \leq \varepsilon\}} d\langle \sigma B \rangle_u = \lim_{\varepsilon \to 0} \frac{\sigma^2}{\varepsilon} \int_0^t 1_{\{S_u - \underline{S}_u \leq \varepsilon\}}\, du \quad \text{a.s.}, \tag{5.7}$$

for all $t \geq 0$. ($\langle M \rangle$ denotes the squared variation of the martingale M.)

Consider first the case $\sigma > 0$. The Lebesgue measure of $\mathscr{Z}$ then vanishes, and we may replace in (5.6) the first two terms by S_t, leaving us with

$$L_t = -\underline{S}_t = \tfrac{1}{2} L_t^0 = \lim_{\varepsilon \to 0} \frac{\sigma^2}{2\varepsilon} \int_0^t 1_{\{S_u - \underline{S}_u \leq \varepsilon\}}\, du.$$

We have thus identified $L = -\underline{S}$ as a local time at 0 of the Markov process $S - \underline{S}$. However, $0 \leq S - \underline{S} \leq \varepsilon$ at a time $t \geq 0$ if and only if $0 \leq \xi_* \leq \varepsilon \cdot 1$ at that time, and ξ_* is Markov, being what came to be known as Walsh's Brownian motion ([23] and [2]). This completes the proof for $\sigma > 0$.

When $\sigma = 0$ and $c < 0$, the process S is of bounded variation. Then L^0 vanishes, the Lebesgue measure of $\mathscr{Z}$ is positive and

$$\underline{S}_t = -c \times \text{Lebesgue measure of } \mathscr{Z} \cap [0, t], \qquad t \geq 0.$$

Again, $-\underline{S}$ is a local time at 0 of $S - \underline{S}$ and ξ_*. Finally, when $\sigma = 0$ and $c \geq 0$, following the leader is achieved by pulling a single arm, and the theorem is trivially true. □

6. Index calculations. The present section is devoted to calculations of some indices and values. Specifically, we analyze symmetric Lévy bandits with no downward jumps, and then bandits driven by compound Poisson, Brownian and Stable processes.

6.1. *No downward jumps.*

6.1.1. *Symmetric bandits.* Consider a symmetric bandit with no downward jumps. Let $X^k(0) = a$, $r^k = r$ and $\Gamma^k = \Gamma$ denote their common ingredients. Then, it follows from (3.5), (5.2), (5.4) and Theorem 5.1 that

$$v(a, \ldots, a) = P_a^k \int_0^\infty e^{-\beta t} \Gamma\left(\frac{1}{d} \underline{X}_t^k\right) dt.$$

In the special case $r(x) = e^{cx}$, $c > 0$,

$$v(a,\dots,a) = \frac{1}{\overline{\varphi}(\beta,-c)\varphi(\beta,c/d)}e^{ca},$$

provided $\overline{\varphi}(\beta,-c)$ makes sense, which is equivalent to (2.1).

6.1.2. *Not a subordinator.* The computations, in this case, were "anticipated" by Fristedt ([7], Corollary 9.9, on page 342) and they are summarized here for reference. Since we deal exclusively with a single arm, the superscript k will be omitted for convenience.

Suppose that X has no downward jumps and it is not a subordinator [$\sigma > 0$ or $c < 0$ in (5.5)]. Then the Lévy exponent ψ is defined over the upper half-plane and, for $\beta > 0$, the equation $\psi(i\eta) = -\beta$ has a unique positive solution, say $\eta(\beta)$. Furthermore, $\eta(\cdot)$ is monotone increasing and $\eta(0)$, defined by continuity, satisfies $\psi[i\eta(0)] = 0$. Finally, the Laplace exponents φ and $\overline{\varphi}$ are given by

$$\varphi(\beta,\gamma) = \gamma + \eta(\beta), \qquad \overline{\varphi}(\beta,\gamma) = \frac{\beta + \psi(i\gamma)}{\eta(\beta) - \gamma}, \qquad \gamma \neq \eta(\beta). \tag{6.8}$$

Calculating ψ and $\eta(\cdot)$ thus yields $\overline{\varphi}(\beta,0)/\overline{\varphi}(\beta,\gamma)$. Inverting this Laplace transform identifies μ, via (4.5), and consequently the index Γ in (3.2). We now carry out this procedure for several Lévy processes.

6.1.3. *Brownian motion.* Let $X_t = ct + \sigma B_t$, $t \geq 0$, where B is a standard Brownian motion. Then

$$\psi(\theta) = -i\theta c + \frac{\theta^2\sigma^2}{2}.$$

Hence $\eta(\beta)$ is the unique positive solution of

$$\eta^2\sigma^2 - 2c\eta - 2\beta = 0,$$

that is,

$$\eta(\beta) = \frac{c + \sqrt{c^2 + 2\beta\sigma^2}}{\sigma^2}.$$

Consequently,

$$\overline{\varphi}(\beta,\gamma) = \frac{\sigma^2}{2}(\gamma + \lambda), \qquad \lambda = \frac{\sqrt{c^2 + 2\beta\sigma^2} - c}{\sigma^2}.$$

Thus,

$$\frac{\overline{\varphi}(\beta,0)}{\overline{\varphi}(\beta,\gamma)} = \frac{\lambda}{\gamma + \lambda},$$

which identifies μ as the exponential distribution with parameter λ. One concludes that the index function is given by

$$\Gamma(x) = \int_0^\infty e^{-y} r\left(x + \frac{y}{\lambda}\right) dy.$$

This result is due to Karatzas [11], who computed it as a special case of diffusion-driven bandits. [The dependence of $\Gamma(x)$ on β, here implicit in $\lambda = \lambda(\beta)$, will be suppressed in later examples as well.]

6.1.4. *Compound Poisson.* Let $X_t = -ct + Y_t$, $t \geq 0$, where $c > 0$ and Y is compound Poisson with rate λ and jumps that are exponentially distributed with parameter α. Then

$$\psi(\theta) = i\theta c + \lambda[1 - B^*(\theta)], \qquad B^*(\theta) = \frac{\alpha}{\alpha - i\theta}.$$

Hence $\eta(\beta)$ is the unique positive solution of

$$\beta - \eta c + \frac{\lambda\eta}{\alpha + \eta} = 0,$$

that is,

$$\eta(\beta) = \frac{\lambda + \beta - \alpha c + \sqrt{(\lambda + \beta - \alpha c)^2 + 4\beta c\alpha}}{2c}.$$

It follows that

$$\overline{\varphi}(\beta, \gamma) = c\left(1 - \frac{\alpha}{\alpha + \gamma}q\right), \quad q = \frac{\alpha c - \lambda - \beta + \sqrt{(\lambda + \beta - \alpha c)^2 + 4\alpha\beta c}}{2\alpha c}.$$

Thus

$$\frac{\overline{\varphi}(\beta, 0)}{\overline{\varphi}(\beta, \gamma)} = p + q\frac{p\alpha}{p\alpha + \gamma}, \qquad p = 1 - q,$$

which identifies μ as the measure with an atom at 0 of size $p = 1 - q$, and with weight q it is the exponential distribution with parameter $p\alpha$. One concludes that

$$\Gamma(x) = pr(x) + q\int_0^\infty e^{-y} r\left(x + \frac{y}{p\alpha}\right) dy.$$

6.1.5. *Stable* 1/2. Let $X_t = -ct + Y_t$, $t \geq 0$, where Y is a stable subordinator of index 1/2. By (9.23) on page 350 of [7],

$$\psi(i\eta) = -\eta c + \sqrt{2\eta}.$$

Hence $\eta(\beta)$ is the unique positive solution of

$$-\eta c + \sqrt{2\eta} = -\beta,$$

that is,

$$\eta(\beta) = \frac{1 + \beta c + \sqrt{1 + 2\beta c}}{c^2}.$$

Substituting $\gamma = \lambda^2/2$ in the expression (6.8) for $\overline{\varphi}(\beta, \gamma)$, we get

$$\overline{\varphi}\left(\beta, \frac{\lambda^2}{2}\right) = \frac{-(\lambda^2/2)c + \lambda + \beta}{\left(1 + \beta c + \sqrt{1 + 2\beta c}\right)/c^2 - \lambda^2/2}$$
$$= \frac{c\left(\lambda - \left(\left(1 - \sqrt{1 + 2\beta c}\right)/c\right)\right)}{\lambda + \left(1 + \sqrt{1 + 2\beta c}\right)/c} = c\frac{c\lambda - 1 + \sqrt{1 + 2\beta c}}{c\lambda + 1 + \sqrt{1 + 2\beta c}},$$

which reveals that

$$\overline{\varphi}(\beta, \gamma) = c\frac{c\sqrt{2\gamma} - 1 + \sqrt{1 + 2\beta c}}{c\sqrt{2\gamma} + 1 + \sqrt{1 + 2\beta c}}.$$

Thus

$$\frac{\overline{\varphi}(\beta, 0)}{\overline{\varphi}(\beta, \gamma)} = p + q\frac{\left(\sqrt{1 + 2\beta c} - 1\right)/c}{\sqrt{2\gamma} + \left(\sqrt{1 + 2\beta c} - 1\right)/c}, \quad p = \frac{\sqrt{1 + 2\beta c} - 1}{\sqrt{1 + 2\beta c} + 1}, \quad q = 1 - p,$$

which identifies μ as the measure with an atom at 0 of size p, and with weight q it has the density g on $(0, \infty)$, given by

$$g(x) = \frac{\sqrt{1 + 2\beta c} - 1}{c\sqrt{2\pi x}}$$
$$\times\left[1 - \frac{\sqrt{1 + 2\beta c} - 1}{c}\sqrt{2\pi x}\exp\left(-\frac{x}{c^2}\left(1 + \beta c - \sqrt{1 + 2\beta c}\right)\right)\right.$$
$$\left.\times\left(1 - \Phi\left(\frac{\sqrt{1 + 2\beta c} - 1}{c}\sqrt{2x}\right)\right)\right],$$

where Φ is the normal distribution function. One concludes that

$$\Gamma(x) = pr(x) + q\int_0^\infty r(x + y)g(y)\,dy.$$

6.2. *Negative jumps.* Let X be a compound Poisson process of rate λ, and double exponential jumps with parameter α. Then

$$\psi(\theta) = \lambda\left(1 - \frac{\alpha^2}{\alpha^2 + \theta^2}\right) = \frac{\lambda\theta^2}{\alpha^2 + \theta^2}.$$

The process X_t is symmetric (isotropic in dimension 1). By (9.14) on page 341 in [7],

$$\overline{\varphi}(\beta, \gamma) = c\exp\left\{\frac{1}{\pi}\int_0^\infty \frac{ln(\beta + \psi(\theta\gamma))}{1 + \theta^2}\,d\theta - \frac{1}{2}\int_0^\infty \frac{e^{-t} - e^{-\beta t}}{t}P\{X_t = 0\}\,dt\right\}.$$

In our case, $P\{X_t = 0\} = e^{-\lambda t}$ and

$$\exp\left\{-\frac{1}{2}\int_0^\infty \frac{(e^{-t} - e^{-\beta t})e^{-\lambda t}}{t}\,dt\right\} = \left(\frac{\lambda + 1}{\beta + \lambda}\right)^{1/2}$$

Furthermore,

$$\begin{aligned}
\frac{1}{\pi}\int_0^\infty \frac{\ln(\beta + \psi(\theta\gamma))}{1 + \theta^2}\,d\theta &= \frac{1}{\pi}\int_0^\infty \frac{\ln(\beta + \lambda\theta^2\gamma^2/(\alpha^2 + \theta^2\gamma^2))}{1 + \theta^2}\,d\theta \\
&= \frac{1}{\pi}\int_0^\infty \frac{\ln(\beta\alpha^2 + \theta^2\gamma^2(\lambda + \beta))}{1 + \theta^2}\,d\theta \\
&\quad - \frac{1}{\pi}\int_0^\infty \frac{\ln(\alpha^2 + \theta^2\gamma^2)}{1 + \theta^2}\,d\theta \\
&= \ln\left(\sqrt{\gamma^2(\lambda + \beta)} + \sqrt{\beta\alpha^2}\right) - \ln\left(\sqrt{\gamma^2} + \sqrt{\alpha^2}\right) \\
&= \ln\frac{\left(\gamma\sqrt{\lambda + \beta} + \alpha\sqrt{\beta}\right)}{\gamma + \alpha},
\end{aligned}$$

so that

$$\overline{\varphi}(\beta, \gamma) = c\frac{\gamma\sqrt{\lambda + \beta} + \alpha\sqrt{\beta}}{\gamma + \alpha}\sqrt{\frac{\lambda + 1}{\beta + \lambda}},$$

and thus

$$\frac{\overline{\varphi}(\beta, 0)}{\overline{\varphi}(\beta, \gamma)} = \frac{\sqrt{\beta}(\gamma + \alpha)}{\gamma\sqrt{\lambda + \beta} + \alpha\sqrt{\beta}} = p + q\frac{p\alpha}{p\alpha + \gamma}, \qquad p = \sqrt{\frac{\beta}{\lambda + \beta}},\ q = 1 - p.$$

We have identified μ as the measure with an atom at 0 of size p, and with weight $q = 1 - p$ it is the exponential distribution with parameter $p\alpha$. Consequently

$$\Gamma(x) = pr(x) + q\int_0^\infty e^{-y} r\left(x + \frac{y}{p\alpha}\right) dy,$$

exactly the form of our previous compound Poisson example in Section 6.1.4, but with different weights. [For both indices, the weight p arises from $l > 0$ in (4.3), and the scaling $p\alpha$ is due to the positive jumps being exponential.]

7. Proofs of index representations. The present section is devoted to proving the equivalence of (4.3) and (4.7). As in the previous section, we suppress the arms' superscript k.

The proof is carried out by discrete approximations. Specifically, introduce the random sequences $X^n = \{X_{l/2^n},\ l = 0, 1, 2, \ldots\}$, indexed by $n \geq 1$. Each X^n is a discrete-time random walk. The value function $v(x, \gamma)$ will be approximated by a corresponding value function for X^n, and similarly for the index Γ. The value and the index functions inherit properties of their approximation (for example, monotonicity and lower semicontinuity). With

those at hand, we identify the form of optimal stopping times for $v(x,\gamma)$, which enables the use of excursion theory, especially the excursion formula (4.2).

Let

$$(7.1)\qquad r_n(x) = 2^n P_x \int_0^{2^{-n}} e^{-\beta t} r[X_t]\,dt = 2^n P_0 \int_0^{2^{-n}} e^{-\beta t} r[X_t + x]\,dt.$$

Each $r_n(x)$ is an increasing function of $x \in \mathbb{R}$. Define the value function associated with X^n by

$$\begin{aligned} v_n(x,\gamma) &= \frac{1-\exp(-\beta/2^n)}{\beta} \sup_{T\in N} P_x \sum_{l=0}^{T-1} \exp\left(-\frac{\beta l}{2^n}\right)\left(r_n(X_{l/2^n}) - \gamma\right) \\ &= \frac{1-\exp(-\beta/2^n)}{\beta/2^n} \sup_{T\in \Pi^n} P_x \int_{t=0}^{T} \exp(-\beta t)(r(X_t) - \gamma)\,dt, \end{aligned}$$

where T runs over all stopping times, with values in $N = \{1,2,\dots\}$ in the first expression and $\Pi^n = \{l/2^n\colon l = 0,1,\dots\}$ in the second.

The sequence of functions $v_n(x,\gamma)$ increases, as $n \uparrow \infty$, to a limit which we denote by $v_\infty(x,\gamma)$. Indeed, $(1 - e^{-\beta/2^n})/(\beta/2^n) \uparrow 1$, and suprema are taken over families of stopping times that increase with n.

PROPOSITION 7.1. $v(x,\gamma) = v_\infty(x,\gamma)$.

PROOF. Obviously $v_\infty(x,\gamma) \le v(x,\gamma)$. To show the reversed inequality, fix $\varepsilon > 0$ and pick a stopping time τ, with respect to $\mathcal{F}$, such that

$$P_x \int_0^{\tau} e^{-\beta t}(r(X_t) - \gamma)\,dt > v(x,\gamma) - \varepsilon.$$

Let

$$\tau^n = \begin{cases} (l+1)/2^n, & \text{if } l/2^n < \tau \le (l+1)/2^n, \\ \infty, & \tau = \infty. \end{cases}$$

Then τ^n is a stopping time in Π^n, $\tau^n \downarrow \tau$ as $n \uparrow \infty$,

$$v_n(x,\gamma) \ge \frac{1-\exp(-\beta/2^n)}{\beta/2^n} P_x \int_0^{\tau^n} \exp(-\beta t)(r(X_t) - \gamma)\,dt$$

and

$$P_x \int_0^{\tau^n} e^{-\beta t} r(X_t)\,dt = P_x \int_0^{\infty} 1_{\{\tau^n > t\}} e^{-\beta t} r(X_t)\,dt.$$

Assumption (2.1) now justifies the use of the dominated convergence theorem, which yields

$$\begin{aligned} &\lim_{n\to\infty} \frac{1-\exp(-\beta/2^n)}{\beta/2^n} P_x \int_0^{\tau^n} \exp(-\beta t)(r(X_t) - \gamma)\,dt \\ &\quad = P_x \int_0^{\tau} \exp(-\beta t)(r(X_t) - \gamma)\,dt, \end{aligned}$$

so that, for all $\varepsilon > 0$,

$$v_\infty(x,\gamma) \geq P_x \int_0^\tau e^{-\beta t}(r(X_t) - \gamma)\, dt \geq v(x,\gamma) - \varepsilon. \qquad \square$$

PROPOSITION 7.2. *The function $x \to v(x,\gamma)$ is increasing and lower semicontinuous.*

PROOF. Monotonicity follows once we prove the result for $v_n(x,\gamma)$, for all n. To this end, note that $\{X_{l/2^n}\colon l = 0,1,\ldots\}$ is a random walk and for such, the optimal stopping time is the hitting time of $A(\gamma) = \{x\colon v_n(x,\gamma) = 0\}$.

Denote by T_A the hitting time of a set A. Suppose that $x < y$ and let $\alpha = \exp(-\beta/2^n)$.

$$\begin{aligned}
v_n(x,\gamma) &= \frac{1-\alpha}{\beta} P_x \sum_{l=0}^{T_{A(\gamma)}-1} \alpha^l \big(r_n(X_{l/2^n}) - \gamma\big) \\
&= \frac{1-\alpha}{\beta} P_0 \sum_{l=0}^{T_{A(\gamma)-x}-1} \alpha^l \big(r_n(x + X_{l/2^n}) - \gamma\big) \\
&\leq \frac{1-\alpha}{\beta} P_0 \sum_{l=0}^{T_{A(\gamma)-x}-1} \alpha^l \big(r_n(y + X_{l/2^n}) - \gamma\big) \\
&= \frac{1-\alpha}{\beta} P_y \sum_{l=0}^{T_{A(\gamma)-x+y}-1} \alpha^l \big(r_n(X_{l/2^n}) - \gamma\big) \\
&\leq v_n(y,\gamma).
\end{aligned}$$

The function $x \to v_n(x,\gamma)$ is continuous since $r(x)$ is continuous (see (6.6) in [16]). From $v(x,\gamma) = \lim_{n\to\infty} v_n(x,\gamma)$ it follows that $v(x,\gamma)$ is lower semicontinuous, as a function of $x \in \mathbb{R}$. $\square$

The last proposition guarantees that $v(x,y)$ is attained by a hitting time of a set that takes the form $(-\infty, b(\gamma))$.

PROOF THAT (4.3) EQUALS (4.7). Using (4.2), in a manner that is similar to the derivation of (3.3) in Section 4.2, one shows that

$$\begin{aligned}
&P_x \int_{[0,\, T_{(-\infty,\, b(\gamma))})} e^{-\beta t}(r(X_t) - \gamma)\, dt \\
&\quad = P_x \int_{[0,\, T_{(-\infty,\, b(\gamma))})} e^{-\beta u} \left[\hat{P}_{X_u} \int_0^R e^{-\beta t}(r(\xi_t) - \gamma)\, dt + l(r(X_u) - \gamma) \right] dL_u,
\end{aligned}$$

where $l = 0$ if the Lebesgue measure of M is equal to 0. Let $\Gamma(x)$ be defined as in (4.3), that is,

$$\Gamma(x) = \frac{lr(x) + \hat{P}_x \int_0^R e^{-\beta t} r(\xi_t)\, dt}{l + \hat{P} \int_0^R e^{-\beta t}\, dt}.$$

Note that since r is continuous and nondecreasing, so is Γ. If for some x, $\gamma > \Gamma(x)$, then P_x a.s., for each $t \in M \cap [0, T_{(-\infty, b(\gamma))})$, $\Gamma(X_t) \leq \Gamma(x) < \gamma$. Since L increases on M only, it follows that, if $b(\gamma) < x$,

$$P_x \int_{[0, T_{(-\infty, b(\gamma))})} e^{-\beta t}(r(X_t) - \gamma)\, dt < 0.$$

Therefore, $b(\gamma) > x$ and $v(x, \gamma) = 0$. If $\gamma < \Gamma(x)$, then there exists $b^* < x$ so that $\Gamma(b^*) > \gamma$ and

$$v(x, \gamma) \geq P_x \int_{[0, T_{(-\infty, b^*)})} e^{-\beta t}(r(X_t) - \gamma)\, dt > 0.$$

It follows that $\Gamma(x) = \inf\{\gamma\colon v(x, \gamma) = 0\}$, which is its definition in (4.7). □

REMARK. Let $\Gamma_n(x) = \inf\{\gamma\colon v_n(x, \gamma) = 0\}$. Since the sequence $v_n(x, \gamma)$ is increasing in n (for fixed γ and x), so is the sequence $\Gamma_n(x)$. Define $\Gamma_\infty(x) = \lim \Gamma_n(x)$. Then clearly $\Gamma_\infty(x) \leq \Gamma(x)$. To show the inequality in the opposite direction, note that for every $\varepsilon > 0$ there exists n so that

$$0 \leq v(x, \Gamma_\infty(x)) \leq v_n(x, \Gamma_\infty(x)) + \varepsilon \leq \varepsilon.$$

Consequently $\Gamma(x) \leq \Gamma_\infty(x)$. It now follows from our theorem and the continuity of $\Gamma(x)$, via Dini's theorem, that the convergence of $\Gamma_n(x)$ to $\Gamma(x)$ is uniform.

8. Proof of optimality. The proof of optimality amounts to proving the following proposition.

PROPOSITION 8.1. *For any strategy T,*

$$\begin{aligned} P_x\, R(T) &= \sum_{k=1}^{d} P_x \int_0^\infty e^{-\beta t} r^k\left[X^k_{T_k}(t)\right] dT_k(t) \\ &\leq \sum_{k=1}^{d} P_x \int_0^\infty e^{-\beta t} \Gamma^k\left[\underline{X}^k_{T_k}(t)\right] dT_k(t), \end{aligned} \tag{8.1}$$

with equality for any index strategy T^*.

Having established the proposition, one notes that the expression in (8.1) is the value of a deteriorating bandit with the decreasing reward processes $\Gamma^k[\underline{X}^k(t)]$, $t \geq 0$, $k = 1, \ldots, d$. For such bandits [14], the optimal strategy is simply to follow the leader among the reward processes, and an index strategy does this. Summing up,

$$
\begin{aligned}
P_x R(T^*) &= \sum_{k=1}^{d} P_x \int_0^\infty e^{-\beta t} \Gamma^k\left[\underline{X}^k_{T^*_k}(t)\right] dT^*_k(t) \\
&= P_x \int_0^\infty e^{-\beta t} \underline{\Gamma}[T^*(t)]\, dt,
\end{aligned} \tag{8.2}
$$

where for $s = (s_1, \ldots, s_d)$, $\underline{\Gamma}(s) = \bigvee_{1 \leq k \leq d} \Gamma^k[\underline{X}^k_{s_k}]$. Theorem 3.1 is thus proved.

PROOF OF PROPOSITION 8.1. The proof is carried out within the framework of Markovian excursion theory [21]. To this end, take X^k to be a right process, specifically $X^k = (\Omega^k, \mathscr{F}^k_\infty, \mathscr{F}^k_t, X^k_t, \theta^k_t, P^k_x)$ is a canonical realization.

An adaptation of some of the proofs of the homogeneity results in [21] (Chapter III, 23.12 and Section 24), up to proper completion and corresponding almost sure statements, gives rise to the following: Let $A = \{A^k_t, t \geq 0\}$ be an increasing right-continuous $\mathscr{F}^k$-adapted process. Then for each fixed $u > 0$, there exists a right-continuous increasing process $\{H^k_v(u, w, w');\ v \geq 0\}$, $(u, w, w') \in \mathbb{R}_+ \times \Omega \times \Omega$, such that:

- For fixed (u, w), $H^k_\cdot(u, w, \cdot)$ is increasing right-continuous and adapted to $\mathscr{F}^k$.
- For fixed (v, w'), $H^k_v(\cdot, \cdot, w')$ is $\mathscr{F}^k$-predictable.
- $A^k_{u+v}(w) = A^k_u(w) + H^k_v(u, w, \theta^k_u w)$, for all u, v, w.

REMARK. The above prevails with each $\mathscr{F}^k$ replaced by $\hat{\mathscr{F}}^k = \{\hat{\mathscr{F}}^k_t, t \geq 0\}$,

$$
\hat{\mathscr{F}}^k_t = \mathscr{F}^k_t \vee \bigvee_{j \neq k} \mathscr{F}^j_\infty, \qquad t \geq 0,
$$

in view of the independence of X^k, $k = 1, \ldots, d$. Hence, we may and will take $X^k = (\Omega, \hat{\mathscr{F}}^k_\infty, \hat{\mathscr{F}}^k_t, X^k_t, \theta^k_t, P_x)$; Ω and P_x are products as defined in our problem formulation (Section 2) and (θ^k_t) is the shift that operates only on the kth coordinate of $w = (w^1, \ldots, w^d)$.

We now prove the inequality, for any strategy $T = (T_1, \ldots, T_d)$. Let

$$
\zeta(u, w) = \inf\{t : T_k(t, w) > u\}
$$

be the right-continuous inverse of T_k. Fix $u \geq 0$ and define

$$
\tilde{\zeta}^u_k(v, w) = H^k_v\left(u - \ , w, \theta^k_u w\right),
$$

where H is given as above, with $A_v^k = \zeta_k(v)$, $v \geq 0$. Suppose, without loss of generality, that $M^1, \ldots, M^l$ are perfect. Then for $k = 1, \ldots, l$,

$$
\begin{aligned}
&P_x \int_0^\infty \exp(-\beta t) r^k\big[X^k_{T_k}(t)\big]\, dT_k(t) \\
&\quad = P_x \int_0^\infty \exp(-\beta \zeta_k(u)) r^k\big[X^k_u\big]\, du \\
&\quad = P_x \sum_{u \in G^k} \exp\big(-\beta \zeta^k(u-)\big) \int_u^{D^k_u} \exp\big(-\beta(\zeta^k(v) - \zeta^k(u-))\big) r^k\big[X^k_v\big]\, dv \\
&\qquad + P_x \int_0^\infty \exp\big(-\beta \zeta^k(u-)\big) r^k\big[X^k_u\big] 1_{M^k}(u)\, du \\
&\quad = P_x \sum_{u \in G^k} \exp\big(-\beta \zeta^k(u-)\big) \int_u^{D^k_u} \exp\big(-\beta \tilde{\zeta}^u_k(v-u)\big) r^k\big[X^k_v\big]\, dv \\
&\qquad + P_x \int_0^\infty \exp\big(-\beta \zeta^k(u-)\big) r^k\big[X^k_u\big] 1_{M^k}(u)\, du \\
&\quad = P_x \int_0^\infty \exp\big(-\beta \zeta^k(u-)\big) \\
&\qquad \times \left(l^k + \hat{P}^k \int_0^{R^k} \exp\big(-\beta \tilde{\zeta}^u_k(v)\big) r^k\big[X^k_u + \xi^k_v\big]\, dv \right) dL^k_u,
\end{aligned}
\tag{8.3}
$$

where the last equality follows from (4.2).

To simplify the presentation, introduce

$$
\Gamma^k_C(x^k) = \frac{l^k + \hat{P}^k \int_0^{R^k} \exp(-\beta C_v) r^k\big(x^k + \xi^k_v\big)\, dv}{l^k + \hat{P}^k \int_0^{R^k} \exp(-\beta C_v)\, dv},
$$

where $C = \{C_v,\ v \geq 0\}$ is $\mathscr{F}^k$-adapted, right-continuous and nondecreasing. Furthermore, recall from excursion theory (Sharpe [21]; Maisonneuve [14]) that

$$
\begin{aligned}
&\hat{P}^k \int_0^{R^k} \exp(-\beta C_v) r^k\big(x^k + \xi^k_v\big)\, dv \\
&\quad = \lim_{\varepsilon \downarrow 0} \hat{P}^k_{x^k} \left(R^k > \varepsilon;\ P_{X^k_t} \int_0^{T_{(-\infty, x^k]}} \exp\big(-\beta \tilde{C}^\varepsilon_v\big) r^k\big(X^k_v\big)\, dv \right),
\end{aligned}
$$

where $\tilde{C}^\varepsilon$ is obtained from C in a manner that resembles the derivation of $\tilde{\zeta}^u_k$ from ζ_k. Finally, if also $C_v \geq v$ for all $v \geq 0$, then $\tilde{C}^\varepsilon_v \geq v$, $v \geq 0$, as well, which by Proposition 4.1 of El-Karoui and Karatzas [5] yields

$$
\Gamma^k_C(X^k) \leq \Gamma^k(X^k). \tag{8.4}
$$

(Proposition 4.1 is a self-contained continuous-time version of Lemma 2.1 in [22].)

We are now ready to complete the proof, taking

$$C_v = \tilde{\zeta}_k^u(v), \qquad v \geq 0.$$

Indeed, multiplying and dividing the integrand in (8.3) by

$$l^k + \hat{P}^k \int_0^{R^k} \exp\bigl(-\beta \tilde{\zeta}_k^u(v)\bigr)\, dv$$

and applying (8.4), one concludes that

$$P_x \int_0^\infty \exp(-\beta t) r^k\bigl[X_{T_k}^k(t)\bigr]\, dT_k(t)$$

$$\leq P_x \int_0^\infty \exp\bigl(-\beta \zeta^k(u-)\bigr)\Gamma^k\bigl[X_u^k\bigr] \times \left(l^k + \hat{P}^k \int_0^{R^k} \exp\bigl(-\beta \hat{\zeta}_k^u(v)\bigr)\, dv\right) dL_u^k \tag{8.5}$$

$$= P_x \int_0^\infty \exp\bigl(-\beta \zeta^k(u)\bigr)\Gamma^k\bigl[\underline{X}_u^k\bigr]\, du \tag{8.6}$$

$$= P_x \int_0^\infty \exp(-\beta t)\Gamma^k\bigl[\underline{X}_{T_k}^k(t)\bigr]\, dT_k(t), \tag{8.7}$$

where (8.6) is obtained from (8.5) through reversing the excursion argument that led to (8.3). Summing up over all $k = 1, \ldots, l$, gives rise to inequality in Proposition 8.1.

For $k = l+1, \ldots, d$, M^k is discrete. This is an easier case because M^k is a countable union of stopping times, all $\mathscr{F}_D$-predictable. The law $\hat{P}^k$ is a (properly normalized) conditional expectation and $l^k = 0$. One may take here, for $u \geq 0$,

$$\tilde{\zeta}_k^u(v, w) = H_v^k\bigl(u, w, \theta_u^k w\bigr).$$

The rest of the argument is almost identical to the above for $k \leq l$, except that one does not need to go through the limiting sequence $\{R^k > \varepsilon\}$ to obtain again

$$P_x \int_0^\infty e^{-\beta t} r^k\bigl[X_{T_k}^k(t)\bigr]\, dT_k(t) \leq P_x \int_0^\infty e^{-\beta t}\Gamma^k\bigl[\underline{X}_{T_k}^k(t)\bigr]\, dT_k(t).$$

Summing up over all $k = 1, \ldots, d$ gives rise to the inequality in Proposition 8.1.

The equality for any index strategy T^* is a consequence of the fact that T^* does not switch arms within interiors of their excursion intervals. Thus, for $k = l+1, \ldots, d$, $\tilde{\zeta}_k^s(v, w) = v - s$, for $s \in G^k$, $v \in (s, D_s^k)$. For $k = 1, \ldots, l$, we now show that almost surely, $\zeta_k(s-) = \zeta_k(s)$ for all $s \in G^k$. This results from "follow the leader" (3.4) in the following manner. When an index strategy is unique, then points in G^k are points of increase of the index and

one would not switch a leading arm. This will be the case if $X^k_{\tau^k}$ is strictly decreasing, for all $k = 1, \ldots, l$; for then, by the quasi left continuity of $(X^k_{\tau^k}, \tau^k)$, different arms do not exit their respective M^k's through identical values of $\underline{X}^k$'s. If, for one or more arms, $X^k_{\tau^k}$ is not strictly decreasing, its sojourn times at points are exponentially distributed. Hence, even if such arms are pulled simultaneously, they almost surely do not start an excursion from their respective M^k's at the same time. Then again, "follow the leader" ensures that $\zeta_k(s -) = \zeta_k(s)$, for $s \in G^k$. It now follows that for $k = 1, \ldots, l$, $\tilde{\zeta}^s_k(v, w) = v - s$ for $s \in G^k$, $v \in (s, D^k_s)$. □

REFERENCES

[1] AZÉMA, J. and YOR, M., eds. (1978). *Temps locaux. Astérisque* **52–53**.

[2] BARLOW, M., PITMAN, J. and YOR, M. (1989). On Walsh's Brownian motion. *Séminaire de Probabilités XXIII. Lecture Notes in Math.* **1372** 275–293. Springer, Berlin.

[3] BERRY, D. A. and FRISTEDT, D. (1985). *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall, London.

[4] CAIROLI, R. and DALANG, R. C. (1994). *Sequential Stochastic Optimization*. To appear.

[5] EL KAROUI, N. and KARATZAS, I. (1994). Dynamic allocation problems in continuous time. *Ann. Appl. Probab.* **4** 255–286.

[6] EL KAROUI, N. and KARATZAS, I. (1994). Synchronization and optimality for general continuous time, multi-armed bandit problems. Preprint.

[7] FRISTEDT, B. (1969). Sample functions of stochastic processes with stationary independent increments. *Adv. in Probab.* **3** 241–396.

[8] GITTINS, J. C. (1989). *Multi-armed Bandit Allocaton Indices*. Wiley, New York.

[9] GITTINS, J. C. and JONES, D. M. (1974). A dynamic allocation index for the sequential design of experiments. In *Progress in Statistics* (J. Gani, K. Sarkadi and I. Vincze, eds.) 241–266. North-Holland, Amsterdam.

[10] JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Berlin.

[11] KARATZAS, I. (1984). Gittins indices in the dynamic allocation problem for diffusion processes. *Ann. Probab.* **12** 173–192.

[12] KASPI, H. and MAISONNEUVE, B. (1984 / 1985). Predictable local times and exit systems. *Seminaire de Probabilities XX. Lecture Notes in Math.* **1204** 95–100. Springer, Berlin.

[13] KASPI, H. and MAISONNEUVE, B. (1988). Regenerative systems on the real line. *Ann. Probab.* **16** 1306–1332.

[14] MAISONNEUVE, B. (1975). Exit systems. *Ann. Probab.* **3** 399–411.

[15] MAISONNEUVE, B. (1983). Ensembles régénératifs de la droite. *Z. Wahrsch. Verw. Gebiete* **63** 501–510.

[16] MANDELBAUM, A. (1986). Discrete multiarmed bandits and multiparameter processes. *Probab. Theory Related Fields* **71** 129–147.

[17] MANDELBAUM, A. (1987). Continuous multi-armed bandits and multi-parameter processes. *Ann. Probab.* **15** 1527–1556.

[18] MANDELBAUM, A., SHEPP, L. A. and VANDERBEI, R. J. (1990). Optimal switching between a pair of Brownian motions. *Ann. Probab.* **18** 1010–1033.

[19] MANDELBAUM, A. and VANDERBEI, R. J. (1993). Brownian bandits. *Temps locaux. Asterisque* **52–53**.

[20] PRESMAN, E. L. and SONIN, I. N. (1990). *Sequential Control with Incomplete Information: The Bayesian Approach to Multi-armed Bandit Problems*. Academic Press, New York.

[21] SHARPE, M. (1988). *General Theory of Markov Processes*. Academic Press, New York.

[22] VARAIYA, P., WALRAND, J. and BUYUKKOC, C. (1985). Extensions of the multi-armed bandit problem. The discounted case. *IEEE Trans. Automat. Control* **AC-30** 426–439.

[23] WALSH, J. B. (1978). A diffusion with discontinuous local time. *Temps locaux. Astérisque* **52–53** 37–45.
[24] WALSH, J. B. (1981). Optional increasing paths. *Colloque ENST-CNET. Lecture Notes in Math.* **863** 172–201. Springer, Berlin.
[25] WEBER, R. (1992). On the Gittins index for multi-armed bandits. *Ann. Appl. Probab.* **2** 1024–1035.
[26] WHITTLE, P. (1980). Multi-armed bandits and the Gittins index. *J. Roy Statist. Soc. Ser. B* **42** 143–149.

FACULTY OF INDUSTRIAL ENGINEERING
AND MANAGEMENT
TECHNION–ISRAEL INSTITUTE OF TECHNOLOGY
TECHNION CITY
HAIFA 32000
ISRAEL