EXPLICIT GITTINS INDICES FOR A
CLASS OF SUPERDIFFUSIVE PROCESSES

ROGER FILLIGER,* *Universität Bielefeld*

MAX-OLIVIER HONGLER,** *Ecole Polytechnique Fédérale de Lausanne*

Abstract

We explicitly calculate the dynamic allocation indices (i.e. the Gittins indices) for multi-armed Bandit processes driven by superdiffusive noise sources. This class of model generalizes former results derived by Karatzas for diffusive processes. In particular, the Gittins indices do, in this soluble class of superdiffusive models, explicitly depend on the noise state.

Keywords: Gittins index; optimal stopping; superdiffusion

2000 Mathematics Subject Classification: Primary 60G40
Secondary 90B36; 93E20

1. Introduction

After the seminal contributions of Gittins [4] and Gittins and Jones [5], the dynamic allocation of a single resource to several competing projects (i.e. the multi-armed Bandit problem) became a classical optimization problem. In its simplest form, this problem considers N independent dynamically evolving projects $j = 1, \ldots, N$, each of which can be in one of many possible states $x_j \in \mathbb{R}$. At each instant of time $t \geq 0$, we are allowed to engage exactly one single project denoted by $i(t)$ and when engaged a project delivers a reward generally dependent on the state $x_i(t)$. In addition, the states of all disengaged projects remain 'frozen' in time. Future earnings are discounted by a factor α and our objective is to schedule, in real time, the engagement of the various projects in order to maximize the expected total discounted reward. For this class of problem, it is possible to define, for each project, a priority index (called the Gittins index) and thus establish an optimal scheduling policy by engaging, at any time, the project exhibiting the highest index. The difficulty is then to calculate the priority indices associated with the dynamics driving the states of the projects. So far, several works have focused on explicitly calculating the priority indices [3], [8], [11], [12]. As calculating a priority index is equivalent to solving an associated optimal stopping problem it is relatively demanding to obtain explicit solutions and, thus, always worthwhile to point out the situations where this can be achieved. It is precisely the aim of this paper to show that explicit Gittins indices can be obtained for projects driven by a class of superdiffusive noise sources. In our models, the noise sources are themselves diffusion processes with nonlinear drifts yielding the superdiffusive behavior, where the variance of the noise grows with time t as $t + Bt^2$ with $B \geq 0$ a constant. For $B = 0$, our class of model has been studied previously by Karatzas [11]; whereas, for $B > 0$,

the superdiffusive dynamics yield noise dependent Gittins indices, a feature not considered in the work of Karatzas [11].

2. The dynamic allocation problem

Following the notation in [11] we denote by $(\Omega, \mathcal{F}, \mathrm{P}; \mathcal{F}_t)$ a standard probability space, where $\{\mathcal{F}_t\}_{t \geq 0}$ is an increasing σ-field in $\mathcal{F}$. We have an N-dimensional Brownian motion $\{w(t) = (w_1(t), \dots, w_N(t)); \mathcal{F}_t, t \geq 0\}$ as well as an allocation policy $\mathcal{A} = \{i(t); \mathcal{F}_t, t \geq 0\}$ which is assumed to be an $\mathcal{F}_t$-progressively measurable process with values in the set $\{1, \dots, N\}$. At time t, the project $i(t) \in \{1, \dots, N\}$ is to be engaged and

$$T_j(t) = \mathrm{meas}\{0 \leq s \leq t \mid i(s) = j\}$$

measures the total time up to t that the project j has been engaged.

The dynamics of the projects are defined on the same probability space and satisfy the following stochastic differential equations:

$$dx_j(t) = \mu_j \, dT_j(t) + dz_j(T_j(t)), \qquad x_j(0) = x_j, \qquad (1)$$

$$dz_j(t) = \sqrt{B} \tanh(\sqrt{B} z_j(t)) \, dt + dw_j(t), \qquad z_j(0) = z, \qquad (2)$$

for $j = 1, \dots, N$. The process $\{x_j(t); t \geq 0\}$ models the 'state of affairs' in project j as an anomalous diffusion with local drift $\mu_j \geq 0$. For ease of presentation, the drift is assumed to be constant. The noise sources $z_j(t)$ show superdiffusive behaviour [9] (i.e. $\mathrm{E}[z_j(t)^2] > t$) as we have

$$\mathrm{E}[z_j^2(t)] = t + Bt^2.$$

The parameter $0 \leq B \leq \min_j\{\mu_j^2\}$ controls the superdiffusive nature of z_j and we observe that for $B = 0$, the dynamics reduce to the known diffusive case treated in [11].

Remark. The specific noise $z_j(t)$ (introduced in [6], [7]) has recently been found to be relevant to the continuous description of a quantum random walk process [2] and is discussed further, together with other soluble superdiffusive models, in [9]. A remarkable representation of the noise z_j has been proposed by Rogers and Pitman [13]. They showed that $z_j(t)$ can be defined as a random mixture of two diffusions of the form

$$z_j(t) = \mathcal{B}_j \sqrt{B} t + w_j(t), \qquad z_j(0) = z, \qquad (3)$$

where $\mathcal{B}_j$ is a Bernoulli variable taking the values $+1$ or -1, independently of the process $w_j(t)$, with respective probabilities $p(z)$ and $1 - p(z)$, where

$$p(z) = \frac{1}{2} + \frac{\tanh(\sqrt{B} z)}{2}.$$

This representation shows, in particular, that with probability 1 we have $\lim_{t \to \infty} z_j(t) = \pm \infty$. Therefore, for large t, (1) and (2) are close to the ordinary diffusions

$$dx_j(t) = (\mu \pm \sqrt{B}) \, dt + dw_j(t), \qquad x_j(0) = x_j,$$

where, however, the sign ± 1 is chosen randomly according to the Bernoulli variable $\mathcal{B}_j$ and the initial condition z_j.

The reward earned by engaging j is measured by the function $h(j, x)$, $x \in \mathbb{R}$, which we assume to be strictly increasing with bounded and continuous first and second derivatives, and

which satisfies

$$\lim_{x \to \infty} h(j, x) = \alpha K, \qquad \lim_{x \to -\infty} h(j, x) = \alpha k, \qquad \lim_{|x| \to \infty} h'(j, x) = 0, \qquad (4)$$

for some numbers $K > k$ and $\alpha > 0$, and for all $j \in \{1, \dots, N\}$. Future earnings are discounted by $e^{-\alpha t}$ and the expected reward corresponding to an initial position $(\boldsymbol{x}, \boldsymbol{z}) = (x_1, \dots, x_N, z_1, \dots, z_N)$ and an allocation policy $\mathcal{A}$ is given by

$$J(\boldsymbol{x}, \boldsymbol{z}; \mathcal{A}) = \mathrm{E}_{\boldsymbol{x}, \boldsymbol{z}} \left[\int_0^\infty e^{-\alpha t} h(i(t), x_{i(t)})(t) \, dt \right]. \qquad (5)$$

The allocation problem is then to find a policy $\mathcal{A}^*$ which maximizes (5).

Following Whittle [14] and Karatzas [11], we treat the optimization problem by adding the extra option of 'retirement', i.e. of abandoning all projects with an associated fixed payoff $m \in [k, K]$. The expected total reward corresponding to an allocation policy $\mathcal{A}$ and a retirement time τ (i.e. a stopping time with respect to $\mathcal{F}_t$) is given by

$$J(\boldsymbol{x}, \boldsymbol{z}, m; \mathcal{A}, \tau) = \mathrm{E}_{\boldsymbol{x}, \boldsymbol{z}} \left[\int_0^\tau e^{-\alpha t} h(i(t), x_{i(t)})(t) \, dt + m e^{-\alpha \tau} \right].$$

The optimization problem is then to choose $\mathcal{A}^*$ and τ^* such that

$$J(\boldsymbol{x}, \boldsymbol{z}, m; \mathcal{A}^*, \tau^*) = \sup_{\mathcal{A}, \tau} J(\boldsymbol{x}, \boldsymbol{z}, m; \mathcal{A}, \tau) \quad \text{for all } m \in \mathbb{R} \text{ and for all } \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^N,$$

where the supremum is taken over all allocation policies and $\mathcal{F}_t$-stopping times τ. This global optimization problem, for which we can write down the characterizing variational inequalities (see, e.g. [11, Equation (2.7)]) is reduced to a stopping time problem using the 'Whittle reduction'. This device transforms the dynamic allocation problem to N optimal stopping problems for the component processes viewed independently of one another. It also distinguishes the form of the optimal allocation policy by assigning an index function $M_j(\cdot)$ to the state of each project and proceeding as follows: 'at time t, engage the project with the leading index $M_j(x_j(t)z_j(t))$ as long as the index exceeds m; otherwise retire'. In the next section we study the generic stopping problem and characterize $\mathcal{A}^*$ and τ^* as explicitly as possible.

3. Optimal stopping

We consider the evolution of a real valued stochastic coordinate X_t defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P}; \mathcal{F}_t)$ by

$$dX_t = \mu \, dt + dZ_t, \qquad\qquad X_0 = x, \qquad (6)$$

$$dZ_t = \sqrt{B} \tanh(\sqrt{B} Z_t) \, dt + dW_t, \qquad Z_0 = z, \qquad (7)$$

where W_t denotes a Brownian motion process adapted to the family $\mathcal{F}_t$ and where B is a positive parameter. The generic *optimal stopping problem* associated to the above dynamic allocation problem reads as follows. Find an $\mathcal{F}_t$-adapted stopping time τ, so as to minimize the expected discounted reward

$$\mathrm{E}_{x, z} \left[\int_0^\tau e^{-\alpha t} h(X_t) \, dt + m e^{-\alpha \tau} \right],$$

where $\alpha > 0$ is a discount factor for the instantaneous reward function $h(x) \in \mathcal{C}_b^1(\mathbb{R})$ which is supposed to be a strictly increasing function satisfying (4).

The general theory [1, Chapter 1, Section 3] established that the optimal stopping time is given by

$$\tau^* = \tau^*(x, z, m) \equiv \inf\{t \geq 0 \mid \phi(x_t, z_t, m) = m\}, \tag{8}$$

where, for fixed m, the $\mathcal{C}_b^1(\mathbb{R}^2)$ function $\phi(x, z, m)$ satisfies the following variational inequalities:

$$\phi(x, z, m) \geq m \quad \text{in } \mathbb{R}^2, \tag{9}$$

$$G\phi(x, z, m) + h(x) \leq \alpha\phi(x, z, m) \quad \text{almost everywhere (a.e.) in } \mathbb{R}^2, \tag{10}$$

$$[G\phi(x, z, m) + h(x) - \alpha\phi(x, z, m)][\phi(x, z, m) - m] = 0 \quad \text{a.e. in } \mathbb{R}^2. \tag{11}$$

Equation (10) is the continuation condition and (11), complementary to (10), expresses the stopping condition. In both equations, G stands for the backward-Kolmogorov operator associated with the project dynamics given in (6) and (7), and is given by

$$G = (\mu + \sqrt{B}\tanh(\sqrt{B}z))\frac{\partial}{\partial x} + \sqrt{B}\tanh(\sqrt{B}z)\frac{\partial}{\partial z} + \frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial x\partial z} + \frac{1}{2}\frac{\partial^2}{\partial z^2}.$$

Let us now show that for the dynamics in (6) and (7), the variational inequality can be solved explicitly. Following Karatzas [11], we first note that the increasing nature of the payoff function h suggests that the continuity region for the stopping problem should be an open interval (b, ∞). In our case, however, b will depend on the state of the driving noise Z_t and we are hence looking for a real boundary point $b = b(Z_t, m)$ and a bounded $\mathcal{C}^1$ function $\phi(x, z, m)$ satisfying the following conditions:

$$\begin{aligned} G\phi(x, z, m) - \alpha\phi(x, z, m) &= -h(x) \quad \text{for } x > b(z, m), \\ G\phi(x, z, m) - \alpha\phi(x, z, m) &< -h(x) \quad \text{for } x < b(z, m), \end{aligned} \tag{12}$$

$$\begin{aligned} \phi(x, z, m) &= m \quad \text{for } x \leq b(z, m), \\ \phi(x, z, m) &> m \quad \text{for } x > b(z, m). \end{aligned} \tag{13}$$

Clearly, (12) and (13) imply (9)–(11). To solve the homogeneous part of (12) (i.e. the first equation with $h = 0$), we remark that the function

$$g(x, z) := \frac{\chi(x, z)}{\cosh(\sqrt{B}z)},$$

solves (12) when χ is a solution to the linear equation

$$\left(\frac{1}{2}\frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial x\partial z} + \frac{1}{2}\frac{\partial^2}{\partial x^2} + \mu\frac{\partial}{\partial x} - \left(\alpha + \frac{B}{2}\right)\right)\chi(x, z) = 0. \tag{14}$$

We have to select a solution to (14) which remains bounded for fixed x and for $|z| \to \infty$. This yields, for the homogeneous part of (12), the following solution:

$$g(x, z) = \frac{1}{2}\frac{e^{\sqrt{B}z}}{\cosh(\sqrt{B}z)}(C_1 e^{x\beta_+} + C_2 e^{-x\gamma_+}) \tag{15}$$

$$+ \frac{1}{2}\frac{e^{-\sqrt{B}z}}{\cosh(\sqrt{B}z)}(C_3 e^{x\beta_-} + C_4 e^{-x\gamma_-}),$$

where $\gamma_\pm = (\mu \pm \sqrt{B}) + r_\pm$, $\beta_\pm = -(\mu \pm \sqrt{B}) + r_\pm$, and $r_\pm = \sqrt{(\mu \pm \sqrt{B})^2 + 2\alpha}$. A particular solution to the nonhomogenous equation (12) is given by the following expected reward of continuing forever:

$$
\begin{aligned}
p(x, z) &= \mathrm{E}_{x,z} \int_0^\infty \mathrm{e}^{-\alpha t} h(X_t) \, \mathrm{d}t \\
&= \sum_{j=0}^1 \frac{\exp[(-1)^j \sqrt{B}z]}{2\cosh(\sqrt{B}z)} \mathrm{E} \int_0^\infty \mathrm{e}^{-\alpha t} h[x + (\mu + (-1)^j \sqrt{B})t + w_t] \, \mathrm{d}t \\
&= \frac{1}{2} \frac{\exp(\sqrt{B}z)}{\cosh(\sqrt{B}z)} \frac{1}{\beta_+ + \gamma_+} \left[\mathrm{e}^{-x\gamma_+} \int_\infty^x \mathrm{e}^{u\gamma_+} h(u) \, \mathrm{d}u + \mathrm{e}^{x\beta_+} \int_x^\infty \mathrm{e}^{-u\beta_+} h(u) \, \mathrm{d}u \right] \\
&\quad + \frac{1}{2} \frac{\exp(-\sqrt{B}z)}{\cosh(\sqrt{B}z)} \frac{1}{\beta_- + \gamma_-} \left[\mathrm{e}^{-x\gamma_-} \int_\infty^x \mathrm{e}^{u\gamma_-} h(u) \, \mathrm{d}u + \mathrm{e}^{x\beta_-} \int_x^\infty \mathrm{e}^{-u\beta_-} h(u) \, \mathrm{d}u \right],
\end{aligned}
$$

where we have used (3) for the conditional probability of the joint process (X_t, Z_t) and standard computations (see, e.g. [10, p. 17]). The general solution to (12) is then given by

$$
\phi(x, z, m) := \phi(x, z) = g(x, z) + p(x, z). \tag{16}
$$

It remains now to determine the constants in (15), which must coincide with the results in [11] for $|z| \to \infty$. We note that the boundedness of ϕ on (b, ∞) implies that $C_1 = C_3 = 0$. Elimination of C_2 and C_4 from

$$
\lim_{z \to \pm\infty} \phi(b(z), z) = m \quad \text{and} \quad \lim_{z \to \pm\infty} \frac{\partial}{\partial x}(\phi(x, z)) \bigg|_{x=b(z)} = 0
$$

yields the asymptotic values $b_\pm = \lim_{z \to \pm\infty} b(z)$ for the switching curve $b(z)$. The asymptotic values $b_\pm$ satisfy the following equations:

$$
m = \frac{p'(b_+, \infty)}{\gamma_+} + p(b_+, \infty) \quad \text{and} \quad m = \frac{p'(b_-, -\infty)}{\gamma_-} + p(b_-, -\infty).
$$

Observe that $b_+ \le b(z) \le b_-$ and that the switching curve $b(z)$ satisfies

$$
\begin{aligned}
m &= \phi(b(z), z) \\
&= g(b(z), z) + p(b(z), z) \\
&= \frac{\mathrm{e}^{\sqrt{B}z}}{2\cosh(\sqrt{B}z)} \left(\frac{p'_+}{\gamma_+} \mathrm{e}^{\gamma_+(\beta_+ + b(z))} + \frac{1}{\beta_+ + \gamma_+} F_+(b(z)) \right) \\
&\quad + \frac{\mathrm{e}^{-\sqrt{B}z}}{2\cosh(\sqrt{B}z)} \left(\frac{p'_-}{\gamma_-} \mathrm{e}^{\gamma_-(\beta_- + b(z))} + \frac{1}{\beta_- + \gamma_-} F_-(b(z)) \right),
\end{aligned} \tag{17}
$$

where

$$
p'_\pm = \lim_{z \to \pm\infty} \frac{\partial}{\partial x} p(x, z) \bigg|_{x=b(z)}
$$

and

$$
F_\pm(b(z)) = \mathrm{e}^{-b(z)\gamma_\pm} \int_{-\infty}^{b(z)} \mathrm{e}^{-u\gamma_\pm} h(u) \, \mathrm{d}u + \mathrm{e}^{b(z)\beta_\pm} \int_{b(z)}^\infty \mathrm{e}^{-u\beta_\pm} h(u) \, \mathrm{d}u.
$$

Note that, for $B = 0$, (17) reduces to Equation (3.14) of [11]. For $B > 0$, (17) defines a switching curve $z \mapsto b(z) = b(z, m)$ which depends on the initial condition z. The uniqueness of this curve is established by fixing $m \in [k, K]$ and $z \in \mathbb{R}$, and by proceeding analogously to Lemma 3.1 of [11]. In particular, we have the following result.

Theorem 1. *For each $m \in [k, K]$ and each initial condition z, there exists a unique number $b(z, m) \in \overline{\mathbb{R}}$ such that the optimal stopping time (8) is given by*

$$\tau^* = \tau(x, z, m)^* = \inf\{t \geq 0 \mid x_t \leq b(z, m)\},$$

where $b(z, m)$ and the total expected reward $\phi(x, z, m)$ are given by (17) and (16), respectively.

4. The Gittins indices

Suppose now that we are given an allocation problem with N projects evolving according to the dynamics given by (1) and (2). The Gittins index $M_j(x_j, z_j)$ for project j is defined as the smallest value of the terminal reward which makes immediate stopping profitable, if the project is in state x_j and the noise in state z_j. The value for M_j follows from solving (17), and the optimal policy $\mathcal{A}^*$ will be 'to continue the project with leading index'.

Acknowledgements

References

[1] BENSOUSSAN, A. AND LIONS, J. L. (1978). *Applications des Inéquations Variationnelles en Contrôle Stochastique*. Dunod, Paris.
[2] BLANCHARD, PH. AND HONGLER, M.-O. (2004). Quantum random walks and piecewise deterministic evolutions. *Phys. Rev. Lett.* **92,** 120601, 4 pp.
[3] DUSONCHET, F. (2003). Dynamic scheduling for production systems operating in random environment. PhD Thesis, EPFL.
[4] GITTINS, J. C. (1979). Bandit processes and dynamic allocation indices (with discussion). *J. R. Statist. Soc. B* **41,** 148–177.
[5] GITTINS, J. C. AND JONES, D. M. (1974). A dynamic allocation index for the sequential design of experiments. In *Progress in Statistics*, ed. J. Gani, North-Holland, Amsterdam, pp. 241–266.
[6] HONGLER M.-O. (1979). Exact solutions of a class of nonlinear Fokker–Planck equations. *Phys. Lett. A* **75,** 3–4.
[7] HONGLER M.-O. (1979). Exact time dependent probability density for a nonlinear non-Markovian stochastic process. *Helv. Phys. Acta* **52,** 280–287.
[8] HONGLER, M.-O. AND DUSONCHET, F. (2001). Optimal stopping and Gittins indices for piecewise deterministic evolution process. *J. Discrete Events Systems* **11,** 235–248.
[9] HONGLER, M.-O., FILLIGER, R. AND BLANCHARD, PH. (2006). Soluble models for dynamics driven by a super-diffusive noise. *Physica A* **370,** 301–315.
[10] ITÔ, K. AND MCKEAN, H. P., JR. (1974). *Diffusion Processes and Their Sample Paths*. Springer, Berlin.
[11] KARATZAS, I. (1984). Gittins indices in the dynamic allocation problem for diffusion processes. *Ann. Prob.* **12,** 173–192.
[12] KASPI, H. AND MANDELBAUM, A. (1998). Multi armed bandits in discrete and continuous time. *Ann. Appl. Prob.* **8,** 1270–1290.
[13] ROGERS, L. C. G. AND PITMAN, J. W. (1981). Markov Functions. *Ann. Prob.* **9,** 573–582.
[14] WHITTLE, P. (1980). Multi armed bandits and the Gittins index. *J. R. Statist. Soc. B* **42,** 143–149.