

PREDICTING THE SUPREMUM: OPTIMALITY OF ‘STOP AT ONCE OR NOT AT ALL’

PIETER C. ALLAART,* *University of North Texas*

Abstract

Let $(X_t)_{0 \leq t \leq T}$ be a one-dimensional stochastic process with independent and stationary increments, either in discrete or continuous time. In this paper we consider the problem of stopping the process (X_t) ‘as close as possible’ to its eventual supremum $M_T := \sup_{0 \leq t \leq T} X_t$, when the reward for stopping at time $\tau \leq T$ is a nonincreasing convex function of $M_T - X_\tau$. Under fairly general conditions on the process (X_t) , it is shown that the optimal stopping time τ takes a trivial form: it is either optimal to stop at time 0 or at time T . For the case of a random walk, the rule $\tau \equiv T$ is optimal if the steps of the walk stochastically dominate their opposites, and the rule $\tau \equiv 0$ is optimal if the reverse relationship holds. An analogous result is proved for Lévy processes with finite Lévy measure. The result is then extended to some processes with nonfinite Lévy measure, including stable processes, CGMY processes, and processes whose jump component is of finite variation.

Keywords: Random walk; Lévy process; optimal prediction; ultimate supremum; stopping time; skew symmetry; convex function

2010 Mathematics Subject Classification: Primary 60G40; 60G50; 60J51
Secondary 60G25

1. Introduction

In recent years there has been a great deal of interest in optimal prediction problems of the form

$$\sup_{\tau \leq T} E[f(M_T - X_\tau)], \quad (1.1)$$

where f is a nonincreasing function, $(X_t)_{t \geq 0}$ is a one-dimensional stochastic process, $T > 0$ is a finite time horizon, and $M_T := \sup\{X_t : 0 \leq t \leq T\}$. The supremum in (1.1) is taken over the set of all stopping times adapted to the process $(X_t)_{t \geq 0}$ for which $P(\tau \leq T) = 1$. For the case of Brownian motion, problem (1.1) has been investigated for several reward functions f , though it is often formulated as a penalty-minimization problem in the form

$$\inf_{\tau \leq T} E[\tilde{f}(M_T - X_\tau)], \quad (1.2)$$

where $\tilde{f} := -f$. For instance, Graversen *et al.* [9] solved (1.2) for standard Brownian motion and $\tilde{f}(x) = x^2$. Their result was generalized to $\tilde{f}(x) = x^\alpha$ for arbitrary $\alpha > 0$ by Pedersen [12], who also considered the function $f = \chi_{[0, \varepsilon]}$ for $\varepsilon > 0$ in (1.1). Du Toit and Peskir [7] were the first to extend these results (for power functions f) to Brownian motion with arbitrary

Received 14 June 2011; revision received 20 March 2012.

* Postal address: Department of Mathematics, University of North Texas, 1155 Union Circle #311430, Denton, TX 76203-5017, USA. Email address: allaart@unt.edu

Supported in part by Japanese GCOE Program G08: ‘Fostering Top Leaders in Mathematics—Broadening the Core and Exploring New Ground’.

drift, which required an entirely new approach. More recently, Shiryaev *et al.* [14] considered problem (1.1) for Brownian motion with drift and $f(x) = e^{-\sigma x}$, where $\sigma > 0$. In that case the problem has the natural interpretation of maximizing the expected ratio of the selling price to the eventual maximum price in the Black–Scholes model for stock price movements. They observed that when the drift parameter lies outside a certain critical interval, the optimal rule τ^* becomes trivial, that is, either $\tau^* \equiv 0$ or $\tau^* \equiv T$. A year later, Du Toit and Peskir [8] managed to prove that the optimal rule is trivial also in the critical interval. More precisely, their result was that $\tau^* \equiv 0$ when the drift is negative, and $\tau^* \equiv T$ when the drift is positive. While this may seem intuitively quite plausible, it is nontrivial to prove. Since the optimal rule changes abruptly from 0 to T as the drift parameter passes through 0, Du Toit and Peskir [8] called it a ‘bang–bang’ stopping rule. They also showed that, for the (seemingly quite similar) problem (1.2) with $\tilde{f}(x) = e^{\sigma x}$, the optimal rule is not of bang–bang form, but transitions from $\tau^* \equiv 0$ to $\tau^* \equiv T$ in a nontrivial way throughout the critical interval.

In the discrete-time setting, an analogous result for a Bernoulli random walk was obtained later the same year by Yam *et al.* [15], using ideas from [8]. Here we put $T = N$, a positive integer, and write X_n instead of X_t , where $\{X_n\}_{0 \leq n \leq N}$ is a simple random walk with parameter p . Yam *et al.* [15] considered both the function $f = \chi_0$, the characteristic function of the set $\{0\}$ (in which case the expectation in (1.1) is just the probability of stopping at the ‘top’ of the random walk), and the function $f(x) = e^{-\sigma x}$, and concluded that in both cases the optimal rule is of bang–bang type. More precisely, the optimal rule is $\tau \equiv N$ when $p > \frac{1}{2}$; $\tau \equiv 0$ when $p < \frac{1}{2}$; or any stopping rule τ satisfying $P(X_\tau = M_\tau \text{ or } \tau = N) = 1$ when $p = \frac{1}{2}$. It is worth noting that the case $f = \chi_0$ had already been considered for general *symmetric* random walks more than 20 years earlier by Hlynka and Sheahan [10].

The results for both discrete and continuous times were recently extended in Allaart [1], where it is shown that the bang–bang principle holds for both a Bernoulli random walk and Brownian motion with drift whenever f is nonincreasing and convex. Equivalently, it holds for problem (1.2) when \tilde{f} is nondecreasing and concave, which is the case, for instance, for the natural penalty function $\tilde{f}(x) = x^\alpha$ with $0 < \alpha \leq 1$. Allaart [1] gave simple sufficient conditions on f for the optimal rules to be unique in the discrete-time case, and necessary and sufficient conditions for the case of Brownian motion.

The aim of the present paper is to extend the result further still, to include more general random walks as well as certain Lévy processes. First, in Section 2, it is shown that the bang–bang principle holds for any random walk whose increments stochastically dominate their opposites, or vice versa (see Theorem 2.1). In Section 3 an analogous result is proved for Lévy processes, first for the case of finite Lévy measure (Theorem 3.1), then for the more general case (Theorem 3.2). This appears to require some notion of drift, and, therefore, it seems necessary to impose some additional conditions pertaining to the ‘small jumps’ of the process. One of these conditions can be omitted in the case when f is continuous and bounded (Theorem 3.3), but the author does not know whether it is needed in the general case. The extra conditions may seem restrictive, but they are satisfied by several commonly studied types of Lévy process, including subordinators, symmetric stable processes, and CGMY processes.

A possible application of this research is in finance. Suppose that you buy a share of stock on the first day of the month, which you must sell some time by the end of the month. Perhaps the stock price follows a random walk in discrete time, and your objective is to maximize the probability of selling the stock at the highest price over the month. In that case, let X_t be the random walk, and let $f = \chi_0$. Or perhaps the stock price follows an exponentiated Lévy process, such as geometric Brownian motion, and your goal is to maximize the expected ratio

of the price at the time you sell to the eventual maximum price. In that case, let X_t be the Lévy process, and put $f(x) = e^{-\sigma x}$, where $\sigma > 0$. In both examples the results of this paper imply, under suitable conditions on the process X_t , that it is either optimal to sell the stock immediately, or to keep it until the last day of the month. In fact, the result for the second example remains valid if one takes as the objective function an arbitrary increasing convex function g of the price ratio, since if $g : (0, \infty) \rightarrow \mathbb{R}$ is increasing and convex, then $f(x) = g(e^{-\sigma x})$ is decreasing and convex.

After starting this work, the author learnt that D. Orlov has also extended the bang–bang principle to certain Lévy processes. Unfortunately, an English version of his paper was not available at the time the present article was nearing completion. In addition, a paper by Bernyk *et al.* [4] appeared in which problem (1.2) was solved for stable Lévy processes of index $\alpha \in (1, 2)$ with no negative jumps, for the penalty function $\tilde{f}(x) = x^p$ with $p > 1$. (We observe that, for this case, $f = -\tilde{f}$ is not convex, so the results of the present paper do not apply; indeed, the optimal rule is nontrivial and its determination requires significant analytical tools.) Some of the preparatory work for this last paper was done in [3].

2. The maximum of a random walk

In this section, let $\{X_n\}_{n=0,1,\dots}$ be a random walk with general steps satisfying a form of skew symmetry as follows: $X_0 \equiv 0$, and, for $n \geq 1$, $X_n = \sum_{k=1}^n \xi_k$, where ξ, ξ_1, ξ_2, \dots are independent, identically distributed (i.i.d.) random variables for which either $\xi \geq_{st} -\xi$ or $\xi \leq_{st} -\xi$. Here, ‘ \geq_{st} ’ denotes the usual stochastic order of random variables, defined by

$$X \geq_{st} Y \iff P(X > t) \geq P(Y > t) \text{ for all } t \in \mathbb{R}.$$

(See Chapter 17 of [11] for a general treatment of the stochastic order.) Let $M_n := \max_{0 \leq k \leq n} X_k$ for $n = 0, 1, \dots, N$, where $N \in \mathbb{N}$ is a finite time horizon. For a nonincreasing function $f : [0, \infty) \rightarrow \mathbb{R}$, consider the optimal stopping problem

$$\sup_{0 \leq \tau \leq N} E[f(M_N - X_\tau)], \tag{2.1}$$

where the supremum is over the set of all stopping times $\tau \leq N$ adapted to the natural filtration $\{\mathcal{F}_n\}_{0 \leq n \leq N}$ of the process $\{X_n\}_{0 \leq n \leq N}$. We note that since f is bounded above, the expectation in (2.1) always exists, though it could take the value $-\infty$.

The above setup includes the Bernoulli random walk with arbitrary parameter $p \in (0, 1)$ as a special case, but is of course much more general.

Theorem 2.1. *Assume that either $\xi \geq_{st} -\xi$ or $\xi \leq_{st} -\xi$, and let $f : [0, \infty) \rightarrow \mathbb{R}$ be non-increasing and convex. Consider problem (2.1).*

- (i) *If $\xi \geq_{st} -\xi$, the rule $\tau \equiv N$ is optimal.*
- (ii) *If $\xi \leq_{st} -\xi$, the rule $\tau \equiv 0$ is optimal.*
- (iii) *If $\xi \stackrel{D}{=} -\xi$, any rule τ satisfying $P(X_\tau = M_\tau \text{ or } \tau = N) = 1$ is optimal.*

Remark 2.1. By the assumption of convexity, f must be continuous on $(0, \infty)$, but it may have a jump discontinuity at $x = 0$. Thus, in particular, Theorem 2.1 covers the important case $f = \chi_0$, the characteristic function of the set $\{0\}$. In that case, the problem comes down to maximizing the probability of stopping at the highest point of the walk, so it can be thought of as a random walk version of the secretary (or best-choice) problem.

Remark 2.2. The condition $\xi \geq_{st} -\xi$ holds for any random variable ξ whose distribution is symmetric about some point $m \geq 0$, as is easy to see. For instance, any normal random variable ξ with a nonnegative mean satisfies $\xi \geq_{st} -\xi$. It follows that Theorem 2.1 applies to all ξ with symmetric distributions.

Example 2.1. An example of a nonsymmetric distribution for which $\xi \geq_{st} -\xi$ is the Gumbel extreme value distribution, with distribution function $F(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$. To see this, let $g(x) = \exp(-e^x) + \exp(-e^{-x})$. Then

$$g'(x) = \exp(-e^x + x)[\exp(e^x - e^{-x} - 2x) - 1].$$

Since it is easy to see (for instance, by using a series expansion) that $e^x - e^{-x} - 2x \geq 0$ for $x \geq 0$, it follows that g is increasing on $[0, \infty)$. Also, since $\lim_{x \rightarrow \infty} g(x) = 1$, this means that $g(x) < 1$ for $x \geq 0$. Hence,

$$1 - F(x) \geq F(-x), \quad x \geq 0.$$

So if $\xi \sim F$ then $\xi \geq_{st} -\xi$.

Example 2.2. The condition $\xi \geq_{st} -\xi$ in Theorem 2.1(i) cannot be replaced by the condition $E(\xi) \geq 0$. For instance, let $P(\xi = 3) = \frac{1}{3} = 1 - P(\xi = -1)$, let $f = \chi_0$, and take $n = 2$. Even though $E(\xi) = \frac{1}{3} > 0$, the optimal rule is easily seen to be $\tau \equiv 0$ rather than $\tau \equiv 2$.

In the case of a Bernoulli random walk, simple sufficient conditions on the function f such that the optimal rules given above be unique are given in [1]. There an example is also given to show that, without the convexity of f , the conclusion of Theorem 2.1 may fail in general.

The proof of Theorem 2.1 uses the following generalization of Lemma 2.1 of [1]. Note that, compared to that lemma, a somewhat different method of proof is needed here.

Lemma 2.1. *Let f be as in Theorem 2.1, and suppose that $\xi \geq_{st} -\xi$. Then*

$$E[f(z \vee M_n - X_n)] \geq E[f(z \vee (M_n - X_n))] \tag{2.2}$$

for all $n \leq N$ and all $z \geq 0$.

Since the statement of Lemma 2.1 involves only expectations, we may construct the random walk on a convenient probability space. Recall first that if $X \geq_{st} Y$ then X and Y can be defined on a common probability space (Ω, \mathcal{F}, P) so that $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$. (See, for instance, [11, Theorem 17.B.1].) Thus, on a sufficiently large probability space, we can construct the random variables ξ_1, \dots, ξ_N together with another set of random variables $\tilde{\xi}_1, \dots, \tilde{\xi}_N$ such that the random vectors $(\xi_1, \tilde{\xi}_1), \dots, (\xi_N, \tilde{\xi}_N)$ are independent, $\tilde{\xi}_i \stackrel{D}{=} -\xi_i$ for each i , and $\xi_i \geq \tilde{\xi}_i$ for each i . Let $\tilde{X}_0 \equiv 0$ and $\tilde{X}_n = \sum_{k=1}^n \tilde{\xi}_k$ for $n = 1, 2, \dots, N$. Finally, define $\tilde{M}_n := \max_{0 \leq k \leq n} \tilde{X}_k$, $n = 0, 1, \dots, N$. Clearly, $X_n \geq \tilde{X}_n$ and $M_n \geq \tilde{M}_n$ for every n .

It is also useful to define

$$Z_n := M_n - X_n \quad \text{and} \quad \tilde{Z}_n := \tilde{M}_n - \tilde{X}_n, \quad n = 0, 1, \dots, N.$$

We easily check that

$$Z_n \leq \tilde{Z}_n, \quad n = 0, 1, \dots, N. \tag{2.3}$$

The key to the proof of Lemma 2.1 is that, for each fixed n ,

$$(M_n - X_n, X_n) \stackrel{D}{=} (\tilde{M}_n, -\tilde{X}_n), \tag{2.4}$$

as follows from an easy time-reversal argument.

Proof of Lemma 2.1. The lemma holds trivially (with equality) when $z = 0$, so assume that $z > 0$. We must first deal separately with the case when $E[f(z \vee M_n - X_n)] = -\infty$. Let $\alpha := [f(z) - f(0)]/z$. Then the convexity of f implies that, for all $u \geq 0$,

$$f(u + z) - f(u) \geq \alpha z. \tag{2.5}$$

Using the algebraic inequality $z \vee m - x \leq z \vee (m - x) + z$ (valid for $z \geq 0$ and $m \geq 0$) and the fact that f is nonincreasing, we obtain

$$f(z \vee m - x) \geq f(z \vee (m - x) + z) \geq f(z \vee (m - x)) + \alpha z,$$

in view of (2.5). Thus, if $E[f(z \vee M_n - X_n)] = -\infty$ then $E[f(z \vee (M_n - X_n))] = -\infty$ as well, and the lemma holds in this case.

Assume for the remainder of the proof that $E[f(z \vee M_n - X_n)] > -\infty$. Since n is fixed, we omit the subscripts and write $M = M_n, X = X_n$, and $Z = Z_n$, and similarly for their tilded counterparts. Let

$$h(z, m, x) := f(z \vee m - x) - f(z \vee (m - x)),$$

so that it is to be shown that

$$E[h(z, M, X)] \geq 0. \tag{2.6}$$

The above expectation exists and is finite, because $\alpha z \leq h(z, m, x) \leq |\alpha|z$. We begin by writing

$$E[h(z, M, X)] = E[h(z, M, X); X > 0] + E[h(z, M, X); X < 0].$$

Using (2.4), we can write the second expectation as

$$E[h(z, M, X); X < 0] = E[h(z, \tilde{M} - \tilde{X}, -\tilde{X}); \tilde{X} > 0]. \tag{2.7}$$

On the other hand, we claim that

$$E[h(z, M, X); X > 0] \geq E[h(z, \tilde{M}, \tilde{X}); \tilde{X} > 0]. \tag{2.8}$$

To see this, note that $h(z, M, X) = 0$ on $\{X > 0, M - X > z\}$, and, hence,

$$\begin{aligned} h(z, M, X) \mathbf{1}(X > 0) &= (f(z \vee M - X) - f(z)) \mathbf{1}(X > 0, M - X \leq z) \\ &= (f(\max\{z - X, Z\}) - f(z)) \mathbf{1}(X > 0, Z \leq z) \\ &\geq (f(\max\{z - X, Z\}) - f(z)) \mathbf{1}(\tilde{X} > 0, \tilde{Z} \leq z) \\ &\geq (f(\max\{z - \tilde{X}, \tilde{Z}\}) - f(z)) \mathbf{1}(\tilde{X} > 0, \tilde{Z} \leq z) \\ &= h(z, \tilde{M}, \tilde{X}) \mathbf{1}(\tilde{X} > 0). \end{aligned}$$

Here the first inequality follows since $\{\tilde{X} > 0, \tilde{Z} \leq z\} \subset \{X > 0, Z \leq z\}$ by (2.3), $\max\{z - X, Z\} \leq z$ on $\{X > 0, Z \leq z\}$, and f is nonincreasing. The second inequality follows since f is nonincreasing and $\max\{z - X, Z\} \leq \max\{z - \tilde{X}, \tilde{Z}\}$.

Combining (2.7) and (2.8), we obtain

$$E[h(z, M, X)] \geq E[h(z, \tilde{M}, \tilde{X}) + h(z, \tilde{M} - \tilde{X}, -\tilde{X}); \tilde{X} > 0]. \tag{2.9}$$

Next, the convexity of f implies that, for all $0 \leq x < y$ and all $d > 0$,

$$f(x) - f(x + d) \geq f(y) - f(y + d), \tag{2.10}$$

as is easily checked. Thus, for $z \geq 0$ and $0 < x \leq m$, we have

$$\begin{aligned} h(z, m, x) + h(z, m - x, -x) &= [f(z \vee m - x) - f(z \vee (m - x))] \\ &\quad + [f(z \vee (m - x) + x) - f(z \vee m)] \\ &= [f(z \vee m - x) - f(z \vee m)] \\ &\quad - [f(z \vee (m - x)) - f(z \vee (m - x) + x)] \\ &\geq 0, \end{aligned}$$

where the inequality follows by (2.10) with $d = x$, since $x > 0$ implies that $z \vee m - x \leq z \vee (m - x)$. This, together with (2.9), yields (2.6).

Corollary 2.1. *Under the hypotheses of Lemma 2.1,*

$$E[f(z \vee M_n - X_n)] \geq E[f(z \vee M_n)] \tag{2.11}$$

for all $n \leq N$ and all $z \geq 0$.

Proof. By (2.4), inequality (2.2) can be expressed alternatively as

$$E[f(z \vee M_n - X_n)] \geq E[f(z \vee \tilde{M}_n)]. \tag{2.12}$$

But, $E[f(z \vee \tilde{M}_n)] \geq E[f(z \vee M_n)]$, since $\tilde{M}_n \leq_{st} M_n$ and f is nonincreasing. Thus, (2.11) follows.

Proof of Theorem 2.1. The main idea in the proof below is essentially due to Du Toit and Peskir [8]; see Yam *et al.* [15] for the discrete-time case.

(i) Suppose first that $\xi_1 \geq_{st} -\xi_1$. Construct the random variables ξ_k, X_k, M_k, Z_k and $\tilde{\xi}_k, \tilde{X}_k, \tilde{M}_k, \tilde{Z}_k$ on a common probability space as in the discussion following the statement of Lemma 2.1. Define the σ -algebras

$$\mathcal{G}_k := \sigma(\{\xi_1, \dots, \xi_k, \tilde{\xi}_1, \dots, \tilde{\xi}_k\}), \quad k = 0, 1, \dots, N.$$

It will be important later in the proof that the increments $X_k - X_j$ and $\tilde{X}_k - \tilde{X}_j$ are independent of \mathcal{G}_j for all $0 \leq j \leq k$. Note further that if the stopping time $\tau \equiv N$ is optimal among the set of all stopping times relative to the filtration $\{\mathcal{G}_k\}$ then it is certainly optimal among the stopping times relative to $\{\mathcal{F}_k\}$. Thus, it is sufficient to show that

$$E[f(M_N - X_\tau)] \leq E[f(M_N - X_N)] \tag{2.13}$$

for any stopping time τ relative to $\{\mathcal{G}_k\}$. Define the functions

$$G(k, z) := E[f(z \vee M_k)], \quad D(k, z) := E[f(z \vee M_k - X_k)],$$

for $z \geq 0$ and $k = 0, 1, \dots, N$. Note that $G(k, z)$ and $D(k, z)$ can possibly take the value $-\infty$. Let $\tau \leq N$ be any stopping time. An easy exercise using the independent and stationary increments of the random walk $\{X_k\}$ leads to

$$E[f(M_N - X_\tau) | \mathcal{G}_\tau] = G(N - \tau, Z_\tau) \tag{2.14}$$

and

$$E[f(M_N - X_N) | \mathcal{G}_\tau] = D(N - \tau, Z_\tau). \tag{2.15}$$

Now Corollary 2.1 says that $D(k, z) \geq G(k, z)$, and, hence,

$$E[f(M_N - X_\tau) \mid \mathcal{G}_\tau] \leq E[f(M_N - X_N) \mid \mathcal{G}_\tau].$$

Taking expectations on both sides gives (2.13), as desired.

(ii) Suppose next that $\xi_1 \leq_{st} -\xi_1$. Apply again the construction following the statement of Lemma 2.1, but this time with $\xi_i \leq \tilde{\xi}_i$ for all i . Observe that all the other relationships between random variables and their tilded counterparts are now reversed as well, i.e.

$$X_k \leq \tilde{X}_k, \quad M_k \leq \tilde{M}_k, \quad Z_k \geq \tilde{Z}_k,$$

for $k = 0, 1, \dots, N$. Define the filtration $\{\mathcal{G}_k\}$ and the function $G(k, z)$ as in the proof of part (i) above, and let

$$\tilde{D}(k, z) := E[f(z \vee \tilde{M}_k - \tilde{X}_k)].$$

In place of (2.12), we now have the inequality

$$E[f(z \vee \tilde{M}_k - \tilde{X}_k)] \geq E[f(z \vee M_k)],$$

or, in other words, $\tilde{D}(k, z) \geq G(k, z)$. Furthermore, the fact that f is nonincreasing implies that $G(k, z)$ is nonincreasing in z , and, therefore,

$$G(N - j, Z_j) \leq G(N - j, \tilde{Z}_j)$$

for each j . By (2.4), $E[f(M_N)] = E[f(\tilde{Z}_N)]$. Putting these facts together, we obtain, for any stopping time τ relative to $\{\mathcal{G}_k\}$, by the same kind of reasoning as in the proof of part (i),

$$\begin{aligned} E[f(M_N - X_\tau)] &= E[G(N - \tau, Z_\tau)] \\ &\leq E[G(N - \tau, \tilde{Z}_\tau)] \\ &\leq E[\tilde{D}(N - \tau, \tilde{Z}_\tau)] \\ &= E[f(\tilde{Z}_N)] \\ &= E[f(M_N)]. \end{aligned} \tag{2.16}$$

Hence, the rule $\tau \equiv 0$ is optimal.

(iii) Suppose finally that $\xi_1 \stackrel{D}{=} -\xi_1$. This is a special case of part (i), so the rule $\tau \equiv N$ is optimal. Now let τ be any stopping time such that, with probability 1, $X_\tau = M_\tau$ or $\tau = N$. Since $G(0, z) = f(z) = D(0, z)$ for all $z \geq 0$ and $G(k, 0) = E[f(M_k)] = E[f(\tilde{Z}_k)] = E[f(Z_k)] = D(k, 0)$ for all k , (2.14) and (2.15) give equality in (2.13). Hence, τ is optimal.

3. The maximum of a Lévy process

A careful study of the proofs in the previous section reveals that the essential property of the random walk is its independent and stationary increments. Furthermore, in order to construct the random walk $\{X_n\}$ and its dual $\{\tilde{X}_n\}$ on a common probability space in such a way that the increments of $\{X_n\}$ uniformly dominate those of $\{\tilde{X}_n\}$ (or vice versa), the step-size distribution had to satisfy a type of skew symmetry. With this in mind, we can now extend the result to a much larger class of stochastic processes.

The general continuous-time analog of a random walk is a *Lévy process*, which is defined as a stochastic process on $[0, \infty)$ with independent and stationary increments which starts at 0 and

is continuous in probability. Following standard practice, we assume also that the process has almost surely right-continuous sample paths with left-hand limits everywhere. If $X = (X_t)_{t \geq 0}$ is a (one-dimensional) Lévy process, it is uniquely determined by the Lévy–Khintchine formula

$$E[e^{iuX_t}] = e^{t\eta(u)},$$

where

$$\eta(u) = i\gamma u - \frac{\sigma^2 u^2}{2} + \int_{\mathbb{R} \setminus \{0\}} [e^{iuy} - 1 - iuy\chi_{(-1,1)}(y)]\nu(dy). \tag{3.1}$$

In this expression, the Lévy measure ν satisfies $\int_{\mathbb{R} \setminus \{0\}} (y^2 \wedge 1)\nu(dy) < \infty$, but ν need not be finite. We say that X is generated by the triplet (γ, σ^2, ν) .

Define the supremum process $M = (M_t)_{t \geq 0}$ by

$$M_t := \sup_{0 \leq s \leq t} X_s, \quad t \geq 0.$$

If ν is finite then X is simply the sum of a Brownian motion with drift and a compound Poisson process, and it is straightforward to adapt the result of the previous section. This is done in Subsection 3.1. If ν is not finite, however, complications arise in attempting to couple the process X with its dual, and some additional conditions appear to be needed to overcome these difficulties. This is made precise in Subsection 3.2. Finally, in Subsection 3.3, we eliminate one of the extra conditions in the case when f is continuous and bounded.

3.1. The case of finite ν

We consider first the case when ν is finite. Then we may set

$$b := \gamma - \int_{0 < |y| < 1} y\nu(dy),$$

and express X_t pathwise in the form

$$X_t = bt + \sigma B_t + \sum_{i=1}^{N(t)} \xi_i, \tag{3.2}$$

where B_t is a standard Brownian motion, ξ_1, ξ_2, \dots are i.i.d. random variables with distribution $\nu/|\nu|$, and $(N(t))_{t \geq 0}$ is a Poisson process with intensity $|\nu|$. In this representation, the Poisson process, the Brownian motion, and the ξ_i s are all independent of one another.

Definition 3.1. Let $X = (X_t)_{t \geq 0}$ be a Lévy process of the form (3.2), with finite Lévy measure ν .

- (i) X is *right skew symmetric (RSS)* if $b \geq 0$ and $\nu((a, \infty)) \geq \nu((-\infty, -a))$ for all $a > 0$.
- (ii) X is *left skew symmetric (LSS)* if $b \leq 0$ and $\nu((a, \infty)) \leq \nu((-\infty, -a))$ for all $a > 0$.
- (iii) X is *symmetric* if $b = 0$ and $\nu((a, \infty)) = \nu((-\infty, -a))$ for all $a > 0$.

Note that the condition regarding ν in the definition of RSS is equivalent to $\xi_1 \geq_{st} -\xi_1$, because if the inequality holds for all $a > 0$, it holds for all $a \in \mathbb{R}$. The following result is the analog of Theorem 2.1 for a Lévy process with finite Lévy measure ν .

Theorem 3.1. *Let $X = (X_t)_{t \geq 0}$ be a Lévy process with finite Lévy measure ν , adapted to a filtration $\{\mathcal{F}_t\}$, such that $X_t - X_s$ is independent of \mathcal{F}_s for all $0 \leq s \leq t$. Assume that X is either RSS or LSS, and let f be as in Theorem 2.1. For fixed $T > 0$, consider the problem*

$$\sup_{0 \leq \tau \leq T} E[f(M_T - X_\tau)], \tag{3.3}$$

where the supremum is over all stopping times τ relative to the filtration $\{\mathcal{F}_t\}$ with $P(\tau \leq T) = 1$.

- (i) *If X is RSS, the rule $\tau \equiv T$ is optimal.*
- (ii) *If X is LSS, the rule $\tau \equiv 0$ is optimal.*
- (iii) *If X is symmetric, any rule τ satisfying $P(X_\tau = M_\tau \text{ or } \tau = T) = 1$ is optimal.*

If $\nu = 0$ then X is a Brownian motion with drift. Thus, the above theorem generalizes recent results of Shiryaev *et al.* [14], Du Toit and Peskir [8, Section 4], and Allaart [1].

Definition 3.2. Let $X = (X_t)_{t \geq 0}$ be a Lévy process. The dual process of X , denoted by \tilde{X} , is a process such that $(\tilde{X}_t)_{t \geq 0} \stackrel{D}{=} (-X_t)_{t \geq 0}$. The dual supremum process, denoted by \tilde{M} , is the process defined by $\tilde{M}_t := \sup_{0 \leq s \leq t} \tilde{X}_s$ for $t \geq 0$.

If X is a Lévy process generated by the triplet (γ, σ^2, ν) then \tilde{X} is a Lévy process with triplet $(-\gamma, \sigma^2, \tilde{\nu})$, where $\tilde{\nu}(A) = \nu(-A)$ for any Borel set $A \subset \mathbb{R}$. Note that if X is RSS then \tilde{X} is LSS and vice versa.

Lemma 3.1. *Let X be any Lévy process. Then, for each fixed $t \geq 0$,*

$$(M_t - X_t, X_t) \stackrel{D}{=} (\tilde{M}_t, -\tilde{X}_t).$$

Proof. This is essentially a known fact. Let $I_t := \inf\{X_s : 0 \leq s \leq t\}$. Then, plainly,

$$(\tilde{M}_t, -\tilde{X}_t) \stackrel{D}{=} (-I_t, X_t).$$

According to Proposition 3 of [5, p. 158],

$$(-I_t, X_t - I_t) \stackrel{D}{=} (M_t - X_t, M_t),$$

which is equivalent to

$$(-I_t, X_t) \stackrel{D}{=} (M_t - X_t, X_t).$$

Thus, the lemma follows.

Proof of Theorem 3.1. Assume for the moment that X is RSS. Recall representation (3.2). On the same probability space on which the process X is defined, we construct the dual \tilde{X} as follows. For each $i \in \mathbb{N}$, we can construct out of ξ_i (using an external randomization if necessary) a random variable $\tilde{\xi}_i$ such that $\tilde{\xi}_i \stackrel{D}{=} -\xi_i$, and $\xi_i \geq \tilde{\xi}_i$ pointwise. Now set

$$\tilde{X}_t := -bt + \sigma B_t + \sum_{i=1}^{N(t)} \tilde{\xi}_i, \quad t \geq 0.$$

Then it is easy to see that $(\tilde{X}_t)_{t \geq 0} \stackrel{D}{=} (-X_t)_{t \geq 0}$, and, moreover, the processes X and \tilde{X} satisfy the property that, for all $0 \leq s < t$ and all $\omega \in \Omega$,

$$X_t(\omega) - X_s(\omega) \geq \tilde{X}_t(\omega) - \tilde{X}_s(\omega). \tag{3.4}$$

For $t \geq 0$, define

$$Z_t := M_t - X_t, \quad \tilde{Z}_t := \tilde{M}_t - \tilde{X}_t.$$

As in Section 2, it follows from (3.4) that

$$M_t \geq \tilde{M}_t \quad \text{and} \quad Z_t \leq \tilde{Z}_t \quad \text{for all } t \geq 0.$$

Using these relationships and Lemma 3.1, we can show in exactly the same way as in the proof of Lemma 2.1 that

$$E[f(z \vee M_t - X_t)] \geq E[f(z \vee (M_t - X_t))]$$

for all $t \geq 0$ and all $z \geq 0$.

Next, for $t \geq 0$, let \mathcal{G}_t be the smallest σ -algebra containing both \mathcal{F}_t and $\sigma(\{\tilde{X}_s : 0 \leq s \leq t\})$. Then $\{\mathcal{G}_t\}_{t \geq 0}$ is a filtration with respect to which both X and \tilde{X} are adapted, and, for each $0 \leq s \leq t$, both $X_t - X_s$ and $\tilde{X}_t - \tilde{X}_s$ are independent of \mathcal{G}_s . The rest of the proof is now the same (modulo subscript notation) as the proof of Theorem 2.1, where the analogs of (2.14) and (2.15) follow, since X , being a Lévy process, obeys the strong Markov property.

Question 3.1. *It is clear that when X is RSS, we have $X_t \geq_{st} \tilde{X}_t$ for all $t \geq 0$. Does the converse of this statement hold?*

3.2. The general case

For a general Lévy process with nonfinite Lévy measure ν , the construction in the previous subsection is no longer possible because the jump times are dense in the time interval $[0, T]$. Here we will use the fact that a general Lévy process on $[0, T]$ can always be obtained as the almost-sure uniform limit of a sequence of processes of the form (3.2). However, in order to ensure that this can be done while preserving the uniform domination of increments (i.e. (3.4)), some extra conditions appear to be needed. Let the Lévy–Khintchine representation of $X = (X(t))_{t \geq 0}$ be given by (3.1). (In what follows, it will be notationally more convenient to write $X(t)$ instead of $X_{t \cdot}$.)

Definition 3.3. We say that X is *balanced in its small jumps (BSJ)* if

$$L := \lim_{\varepsilon \downarrow 0} \int_{\varepsilon \leq |y| < 1} y \nu(dy) \quad \text{exists and is finite.} \tag{3.5}$$

This condition is always satisfied when ν is symmetric on a sufficiently small interval $(-\varepsilon, \varepsilon)$, where $\varepsilon > 0$, or when $\int_{0 < |y| < 1} |y| \nu(dy) < \infty$. (In the latter case, the non-Gaussian part of X has finite variation.) In the case when $\int_{0 < |y| < 1} |y| \nu(dy) = \infty$, (3.5) may be interpreted as saying that ν is *almost* symmetric in a sufficiently small neighborhood of the origin. Roughly speaking, this means that we allow the small jumps of the process to be dense in time, provided that the positive and negative jumps more or less balance each other. It allows us to still think of the number $\gamma - L$ as the ‘drift’ of the process.

It is clear that if X is BSJ then so is its dual \tilde{X} .

Denote by $\tilde{\nu}$ the dual measure of ν , so that $\tilde{\nu}(A) = \nu(-A)$ for $A \subset \mathbb{R}$. If μ and ν are measures on \mathbb{R} and $E \subset \mathbb{R}$, we say that μ *majorizes* ν on E if $\mu(F) \geq \nu(F)$ for every $F \subset E$.

Definition 3.4. Let $X = (X(t))_{t \geq 0}$ be a Lévy process.

(i) X is *strongly right skew symmetric (SRSS)* if all of the following statements hold:

- (a) X is BSJ;
- (b) $\gamma \geq L$, where L is the limit in (3.5);
- (c) $\nu((a, \infty)) \geq \nu((-\infty, -a))$ for all $a > 0$;
- (d) there exists $\varepsilon > 0$ such that ν majorizes $\tilde{\nu}$ on $(0, \varepsilon)$.

(ii) X is *strongly left skew symmetric (SLSS)* if \tilde{X} is SRSS.

(iii) X is *symmetric* if $\gamma = 0$ and $\nu = \tilde{\nu}$.

Remark 3.1. (i) If X is symmetric then it is both SRSS and SLSS, since (3.5) holds with $L = 0$.

(ii) If X is SRSS or SLSS, and ν is finite, then X is RSS or, respectively, LSS, since $b = \gamma - L$. The undesirable fourth condition in the definition of SRSS seems to be needed in order to carry out the pathwise construction of X and its dual, below. At this point, the author does not see how to get around this technical difficulty, except in the special case when f is bounded and continuous (see Subsection 3.3).

(iii) The SRSS and SLSS conditions can be made more concrete in the case when ν has a density. Let $f, g: (0, \infty) \rightarrow [0, \infty)$, and suppose that

$$\nu(dx) = (f(x)\chi_{(0,\infty)}(x) + g(-x)\chi_{(-\infty,0)}(x)) dx.$$

Then ν is a Lévy measure if and only if $\int_0^\infty (x^2 \wedge 1)[f(x) + g(x)] dx < \infty$. The BSJ condition is now equivalent to convergence of the integral $\int_0^1 x[f(x) - g(x)] dx$. Conditions (c) and (d) in the definition of SRSS respectively become

- (c') $\int_a^\infty [f(x) - g(x)] dx \geq 0$ for all $a > 0$;
- (d') there exists $\varepsilon > 0$ such that $f(x) \geq g(x)$ for all $x \in (0, \varepsilon)$.

The easiest way to satisfy both (c') and (d') is, of course, to take $f \geq g$ everywhere. This way, we may obtain nontrivial examples of nonsymmetric Lévy processes that are SRSS (or SLSS). For instance, let

$$f(x) = \frac{c}{x^p}, \quad g(x) = \frac{c}{x^p + x^r}, \quad \text{where } c > 0, 2 \leq p < 3, r > 2p - 2.$$

Then $r > p$, and ν satisfies (3.5) with

$$L = \int_0^1 x[f(x) - g(x)] dx = \int_0^1 \frac{cx^{r-2p+1}}{1 + x^{r-p}} dx,$$

a convergent integral. Since (c') and (d') are obviously satisfied, the process will be SRSS if $\gamma \geq L$. Since $r > p$, the ‘large’ positive jumps of $X(t)$ tend to be greater in magnitude (and occur more frequently) than the ‘large’ negative jumps. On the other hand, the small jumps of the process in either direction are comparable in size.

Example 3.1. (*Stable processes.*) Let X be a stable Lévy process with index of stability α ($0 < \alpha \leq 2$). If $\alpha = 2$ then X is just a Brownian motion with drift, and the optimal rule is already specified by Theorem 3.1. (In fact, in this case the optimal rules are unique except for some trivial cases; see [1].)

If $\alpha < 2$ then $\sigma = 0$ and the Lévy measure ν is of the form

$$\nu(dx) = \left(\frac{c_1}{x^{1+\alpha}} \chi_{(0,\infty)}(x) + \frac{c_2}{|x|^{1+\alpha}} \chi_{(-\infty,0)}(x) \right) dx,$$

where $c_1 \geq 0$, $c_2 \geq 0$, and $c_1 + c_2 > 0$ (see, e.g. [13, p. 80]). It follows that if $1 \leq \alpha < 2$ then X is BSJ if and only if $c_1 = c_2$, in which case ν is symmetric. In that case, X is SRSS if $\gamma \geq 0$, and X is SLSS if $\gamma \leq 0$. On the other hand, if $0 < \alpha < 1$ then the BSJ condition (3.5) is always satisfied, with

$$L = \int_{0 < |x| < 1} x \nu(dx) = \frac{c_1 - c_2}{1 - \alpha},$$

and X is SRSS if $\gamma \geq L \geq 0$, or, similarly, X is SLSS if $\gamma \leq L \leq 0$.

Note that in the stable case, condition (d) of Definition 3.4(i) is satisfied whenever conditions (a)–(c) are.

Example 3.2. (*CGMY processes.*) Another example of nonsymmetric processes that are SRSS or SLSS is given by the CGMY processes, which are frequently used in financial modeling. The CGMY process, named after Carr, Geman, Madan and Yor (see [6]), is a Lévy process with Lévy measure

$$\nu(dx) = C \frac{e^{-G|x|} \chi_{(-\infty,0)}(x) + e^{-Mx} \chi_{(0,\infty)}(x)}{|x|^{1+Y}},$$

where $C > 0$, $G \geq 0$, $M \geq 0$, and $Y < 2$, and it is assumed that $G > 0$ and $M > 0$ if $Y \leq 0$. The CGMY processes include the symmetric stable processes (take $G = M = 0$) and are sometimes called *tempered stable processes*. The CGMY process with $Y = 0$ is known as the *variance gamma process*. The very small jumps of a CGMY process behave essentially as in the symmetric stable case, and it is easy to check that all CGMY processes have the BSJ property. Furthermore, conditions (c) and (d) of Definition 3.4(i) are satisfied if and only if $M \leq G$. Hence, the CGMY process is SRSS if $M \leq G$ and $\gamma \geq L$, with L as in (3.5), and it is SLSS if $M \geq G$ and $\gamma \leq L$.

We can now state the result for the most general case.

Theorem 3.2. *Let $X = (X(t))_{t \geq 0}$ be a Lévy process, and let f be as in Theorem 2.1. For fixed $T > 0$, consider problem (3.3).*

- (i) *If X is SRSS, the rule $\tau \equiv T$ is optimal.*
- (ii) *If X is SLSS, the rule $\tau \equiv 0$ is optimal.*
- (iii) *If X is symmetric, any rule τ satisfying $P(X(\tau) = M(\tau) \text{ or } \tau = T) = 1$ is optimal.*

The proof of Theorem 3.2 hinges on the following construction. Once this is accomplished, the rest of the proof is the same as before.

Lemma 3.2. *Let X be an SRSS Lévy process. Then, on a suitable probability space (Ω, \mathcal{F}, P) , we can construct X and its dual \tilde{X} in such a way that there exists a set $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$*

such that, for all $0 \leq s < t$ and all $\omega \in \Omega_0$,

$$X(t; \omega) - X(s; \omega) \geq \tilde{X}(t; \omega) - \tilde{X}(s; \omega). \tag{3.6}$$

Proof. Let $\varepsilon > 0$ be as in the definition of SRSS. Then $X(t)$ can be expressed by the Lévy–Itô decomposition

$$X(t) = \gamma' t + \sigma B(t) + \int_{|y| < \varepsilon} y N'(t, dy) + \int_{|y| \geq \varepsilon} y N(t, dy),$$

where $B(t)$ is a standard Brownian motion on \mathbb{R} , $\gamma' := \gamma - \int_{\varepsilon \leq |y| < 1} y \nu(dy)$, $(N(t, \cdot))_{t \geq 0}$ is a Poisson random measure with intensity measure ν which is independent of the Brownian motion, and $N'(t, \cdot)$ is defined by

$$N'(t, dy) = N(t, dy) - t \nu(dy), \quad t \geq 0.$$

In general, the integrals $\int_{|y| < \varepsilon} y N(t, dy)$ and $\int_{|y| < \varepsilon} y \nu(dy)$ need not converge, but the ‘compensated sum of small jumps’, $\int_{|y| < \varepsilon} y N'(t, dy)$, always does.

Now we will construct a sequence of Lévy processes Y_1, Y_2, \dots and their duals $\tilde{Y}_1, \tilde{Y}_2, \dots$, as follows. Let $\varepsilon = \varepsilon_1 > \varepsilon_2 > \dots$ be a sequence of numbers decreasing to 0. Define first

$$Y_1(t) := (\gamma - L)t + \sigma B(t) + \int_{|y| \geq \varepsilon} y N(t, dy).$$

Then Y_1 has finite Lévy measure ν_1 , where ν_1 is the restriction of ν to the set $\{y : |y| \geq \varepsilon\}$. Clearly, $\nu_1((a, \infty)) \geq \nu_1((-\infty, -a))$ for all $a > 0$, since ν_1 simply inherits this property from ν . Since $\gamma \geq L$, we can construct Y_1 and its dual \tilde{Y}_1 on the same probability space so that these processes satisfy the increment property (3.4). Next, for $n \geq 2$, let

$$Y_n(t) = \int_{\varepsilon_n \leq |y| < \varepsilon_{n-1}} y N(t, dy),$$

and note that, by the usual independence property of Poisson point processes, the processes $Y_n, n \in \mathbb{N}$, may be constructed independently of each other. Now, for each $n \geq 2, Y_n$ is a compound Poisson process with (finite) Lévy measure ν_n , where ν_n is the restriction of ν to the set $\{y : \varepsilon_n \leq |y| < \varepsilon_{n-1}\}$. Since ν majorizes $\tilde{\nu}$ on $(0, \varepsilon)$, it follows that $\nu_n((a, \infty)) \geq \tilde{\nu}_n((a, \infty))$ for all $n \geq 2$. (Note that this fact would not be guaranteed without the fourth condition in the definition of SRSS.) Thus, we can construct Y_n and its dual \tilde{Y}_n together as in the previous subsection in such a way that these processes satisfy (3.4).

Finally, set

$$X_n(t) := Y_1(t) + \dots + Y_n(t), \quad \tilde{X}_n(t) := \tilde{Y}_1(t) + \dots + \tilde{Y}_n(t),$$

for $n \in \mathbb{N}$, so that \tilde{X}_n is the dual of X_n . Since property (3.4) is clearly preserved under addition of two or more processes, we have, for all $0 \leq s < t$,

$$X_n(t) - X_n(s) \geq \tilde{X}_n(t) - \tilde{X}_n(s) \tag{3.7}$$

pointwise on Ω . Finally, note that $X_n(t)$ can be written as

$$\begin{aligned} X_n(t) &= (\gamma - L)t + \sigma B(t) + \int_{|y| \geq \varepsilon_n} y N(t, dy) \\ &= \gamma_n t + \sigma B(t) + \int_{\varepsilon_n \leq |y| < \varepsilon} y N'(t, dy) + \int_{|y| \geq \varepsilon} y N(t, dy), \end{aligned}$$

where

$$\gamma_n := \gamma - L + \int_{\varepsilon_n \leq |y| < \varepsilon} y \nu(dy).$$

By (3.5), $\gamma_n \rightarrow \gamma'$, and it follows from Theorem 2.6.2 of [2] that $X_n(t) \rightarrow X(t)$ uniformly in $[0, T]$ with probability 1, as long as the sequence $\{\varepsilon_n\}$ decreases fast enough so that

$$\int_{0 < |y| < \varepsilon_n} y^2 \nu(dy) \leq \frac{1}{8^n} \tag{3.8}$$

for every n . Similarly, $\tilde{X}_n(t) \rightarrow \tilde{X}(t)$ uniformly in $[0, T]$ with probability 1. By taking limits in (3.7), we see that X and \tilde{X} satisfy (3.6) everywhere on the set on which both processes converge.

3.3. The case of bounded and continuous f

In general, it seems difficult to eliminate the unnatural condition (d) in the definition of SRSS, except when the reward function f is bounded and continuous on $[0, \infty)$. This case includes, for instance, the natural reward function $f(x) = e^{-\sigma x}$ with $\sigma > 0$.

Say a general Lévy process $X = (X(t))_{t \geq 0}$ with Lévy–Khintchine representation (3.1) is RSS if

$$\gamma \geq \liminf_{\delta \downarrow 0} \int_{\delta < |y| < 1} y \nu(dy), \tag{3.9}$$

and $\nu((a, \infty)) \geq \nu((-\infty, a))$ for all $a > 0$. Say X is LSS if \tilde{X} is RSS.

Theorem 3.3. *Let $X = (X(t))_{t \geq 0}$ be a Lévy process, and let $f: [0, \infty) \rightarrow \mathbb{R}$ be bounded, nonincreasing, continuous, and convex. For fixed $T > 0$, consider problem (3.3).*

- (i) *If X is RSS, the rule $\tau \equiv T$ is optimal.*
- (ii) *If X is LSS, the rule $\tau \equiv 0$ is optimal.*

(Observe that the symmetric case is already covered by Theorem 3.2.)

Proof of Theorem 3.3. Suppose first that X is RSS. Let $L := \liminf_{\delta \downarrow 0} \int_{\delta < |y| < 1} y \nu(dy)$, and choose a sequence $\delta_1 > \delta_2 > \dots > 0$ so that $\lim_{k \rightarrow \infty} \int_{\delta_k < |y| < 1} y \nu(dy) = L$. For each n , choose k_n so that $\varepsilon_n := \delta_{k_n}$ satisfies (3.8). Now we construct the process X as an almost-sure uniform limit of a sequence of processes $X_n = (X_n(t))_{t \geq 0}$, $n \in \mathbb{N}$, exactly as in the proof of Lemma 3.2. Then each X_n is RSS in the sense of Subsection 3.1. (Note that in order to construct the processes X_n in this way, without their duals, condition (d) of Definition 3.4(i) is not needed.) For each $t \geq 0$, let \mathcal{F}_t be the smallest σ -algebra containing each $\sigma(\{X_n(s) : 0 \leq s \leq t\})$, $n \in \mathbb{N}$. Let Ω_0 be the subset of Ω on which $X_n(t)$ converges uniformly in t . By arbitrarily redefining $X(t; \omega) \equiv 0$ for $\omega \in \Omega \setminus \Omega_0$, we see that X is adapted to $\{\mathcal{F}_t\}$, and, clearly, $X_t - X_s$ is independent of \mathcal{F}_s for each $0 \leq s \leq t$. Thus, by Theorem 3.1, for any stopping time τ relative to $\{\mathcal{F}_t\}$,

$$E[f(M_n(T) - X_n(\tau))] \leq E[f(M_n(T) - X_n(T))]. \tag{3.10}$$

Now it follows from the uniform convergence of X_n to X that, pointwise on Ω_0 , $M_n(T) \rightarrow M(T)$ and $X_n(\tau) \rightarrow X(\tau)$. Hence, by the continuity of f ,

$$f(M_n(T) - X_n(\tau)) \rightarrow f(M(T) - X(\tau))$$

and

$$f(M_n(T) - X_n(T)) \rightarrow f(M(T) - X(T)).$$

Thus, taking limits in (3.10) we see via the bounded convergence theorem that

$$\mathbb{E}[f(M(T) - X(\tau))] \leq \mathbb{E}[f(M(T) - X(T))].$$

Therefore, the rule $\tau \equiv T$ is optimal. A similar argument shows that the rule $\tau \equiv 0$ is optimal if X is LSS.

Remark 3.2. If we try to extend the above reasoning to unbounded continuous f via the dominated convergence theorem, we run into the difficulty of bounding expectations such as $\mathbb{E}|f(M_n(T))|$ uniformly in n , since there is no guarantee that $\mathbb{E}|f(M_n(T))|$ converges to $\mathbb{E}|f(M(T))|$.

Remark 3.3. It may seem that in Theorem 3.2 we could have weakened the SRSS condition similarly, replacing (a) and (b) in Definition 3.4(i) with (3.9). But, this would not actually give a weaker hypothesis, since, in the presence of condition (d), the integral in (3.5) increases monotonically as $\varepsilon \downarrow 0$.

Acknowledgements

This work was started while the author was on sabbatical in Kyoto, Japan. The author wishes to thank the Kyoto University Mathematics Department and the Research Institute for Mathematical Sciences for their warm hospitality during 2009. The author is grateful to two anonymous referees for several suggestions to improve the presentation of this paper.

References

- [1] ALLAART, P. C. (2010). A general ‘bang–bang’ principle for predicting the maximum of a random walk. *J. Appl. Prob.* **47**, 1072–1083.
- [2] APPLEBAUM, D. (2009). *Lévy Processes and Stochastic Calculus*, 2nd edn. Cambridge University Press.
- [3] BERNYK, V., DALANG, R. C. AND PESKIR, G. (2008). The law of the supremum of a stable Lévy process with no negative jumps. *Ann. Prob.* **36**, 1777–1789.
- [4] BERNYK, V., DALANG, R. C. AND PESKIR, G. (2011). Predicting the ultimate supremum of a stable Lévy process with no negative jumps. *Ann. Prob.* **39**, 2385–2423.
- [5] BERTOIN, J. (1996). *Lévy Processes*. Cambridge University Press.
- [6] CARR, P., GEMAN, H., MADAN, D. B AND YOR, M. (2002). The fine structure of asset returns: an empirical investigation. *J. Business* **75**, 305–332.
- [7] DU TOIT, J. AND PESKIR, G. (2007). The trap of complacency in predicting the maximum. *Ann. Prob.* **35**, 340–365.
- [8] DU TOIT, J. AND PESKIR, G. (2009). Selling a stock at the ultimate maximum. *Ann. Appl. Prob.* **19**, 983–1014.
- [9] GRAVERSEN, S. E., PESKIR, G. AND SHIRYAEV, A. N. (2000). Stopping Brownian motion without anticipation as close as possible to its ultimate maximum. *Theory Prob. Appl.* **45**, 41–50.
- [10] HLYNKA, M. AND SHEAHAN, J. N. (1988). The secretary problem for a random walk. *Stoch. Process. Appl.* **28**, 317–325.
- [11] MARSHALL, A. W. AND OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.
- [12] PEDERSEN, J. L. (2003). Optimal prediction of the ultimate maximum of Brownian motion. *Stoch. Stoch. Reports* **75**, 205–219.
- [13] SATO, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.
- [14] SHIRYAEV, A., XU, Z. AND ZHOU, X. Y. (2008). Thou shalt buy and hold. *Quant. Finance* **8**, 765–776.
- [15] YAM, S. C. P., YUNG, S. P. AND ZHOU, W. (2009). Two rationales behind ‘buy-and-hold or sell-at-once’ strategy. *J. Appl. Prob.* **46**, 651–668.