

# The coding complexity of Lévy processes

by

Frank Aurzada and Steffen Dereich

*Institut für Mathematik, MA 7-5, Fakultät II  
Technische Universität Berlin  
Straße des 17. Juni 136  
10623 Berlin  
{aurzada,dereich}@math.tu-berlin.de*

November 21, 2018

**Summary.** We investigate the high resolution coding problem for general real-valued Lévy processes under  $L^p[0, 1]$ -norm distortion. Tight asymptotic formulas are found under mild regularity assumptions.

**Keywords.** High-resolution quantization; distortion-rate function; complexity; Lévy processes.

**2000 Mathematics Subject Classification.** 60G35, 41A25, 94A15.

## 1 Introduction and Results

### 1.1 Motivation and Notation

In this article, we study the coding problem for real-valued Lévy processes  $X$  (*original*) under  $L^p[0, 1]$ -norm distortion for some fixed  $p \in [1, \infty)$ . Here we think of  $X$  being a  $\mathbb{D}[0, \infty)$ -valued process, where  $\mathbb{D}[0, \infty)$  denotes the space of càdlàg functions endowed with the Skorohod topology. We shall denote by  $\|\cdot\|$  the standard  $L^p[0, 1]$ -norm.

Let  $0 < s \leq \infty$ . The objective is now to find a càdlàg real-valued process  $\hat{X}$  (*reconstruction* or *approximation*) that minimizes the error criterion

$$\| \|X - \hat{X}\| \|_{L^s(\mathbb{P})} = \begin{cases} \mathbb{E}[\|X - \hat{X}\|^s]^{1/s} & \text{if } s < \infty \\ \text{ess sup } \|X - \hat{X}\| & \text{if } s = \infty \end{cases} \quad (1)$$

under a given complexity constraint on the approximating random variable  $\hat{X}$ . We will work with the following three complexity constraints that have been originally suggested by Kolmogorov [6]: for  $r \geq 0$ ,

- $\log |\text{range}(\hat{X})| \leq r$  (*quantization constraint*)
- $H(\hat{X}) \leq r$ , where  $H$  denotes the entropy of  $\hat{X}$  (*entropy constraint*)
- $I(X; \hat{X}) \leq r$ , where  $I$  denotes the Shannon mutual information of  $X$  and  $\hat{X}$  (*mutual information constraint*).

We will work with the following standard notation for entropy and mutual information:

$$H(\hat{X}) = \begin{cases} -\sum_x p_x \log p_x & \text{if } \hat{X} \text{ is discrete with probability weights } (p_x) \\ \infty & \text{otherwise} \end{cases}$$

and

$$I(X, \hat{X}) = \begin{cases} \int \frac{d\mathbb{P}_{X, \hat{X}}}{d\mathbb{P}_X \otimes d\mathbb{P}_{\hat{X}}} d\mathbb{P}_{X, \hat{X}} & \text{if } \mathbb{P}_{X, \hat{X}} \ll \mathbb{P}_X \otimes \mathbb{P}_{\hat{X}} \\ \infty & \text{otherwise.} \end{cases}$$

Here,  $\mathbb{P}_Z$  denotes the distribution function of a random variable  $Z$ .

When considering the quantization constraint, we get the following minimal value

$$D^{(q)}(r, s) := \inf\{\|X - \hat{X}\|_{L^s(\mathbb{P})} : \log |\text{range}(\hat{X})| \leq r\},$$

which we call the (minimal) *quantization error* for the *rate*  $r \geq 0$  and the *moment*  $s$ . Analogously, we denote by  $D^{(e)}(r, s)$  and  $D(r, s)$  the minimal values under the entropy- and mutual information constraint, respectively.  $D^{(e)}$  and  $D$  will be called entropy coding error and distortion rate function, respectively. We have  $D \leq D^{(e)} \leq D^{(q)}$ , for any random variable.

The quantization constraint naturally appears, when coding the signal  $X$  under a strict bit-length constraint. The entropy constraint corresponds to an average bit-length constraint and the mutual information constraint gains its importance from Shannon's celebrated source coding theorem. In this article we will not consider the run time behaviour of our coding schemes. However, we think that the approximation schemes (provided later in the article) have implementations with reasonable runtime behaviour. Strictly speaking, the quantities  $D^{(e)}$  and  $D$  depend on the probability space. However, this dependence has no effect on our results.

The objective of the article is

- to provide efficient coding strategies for general Lévy processes that are parameterized by three parameters and that are robust under a mismatch on the Lévy measure and
- to complement the estimates by appropriate lower bounds that show weak optimality of our scheme for most cases.

In the article,  $X = (X_t)_{t \in [0, \infty)}$  denotes a Lévy process in the Skorohod space  $\mathbb{D}[0, \infty)$ , that is a process starting in 0 with independent and stationary increments. Due to the Lévy Khintchine formula, the characteristic function of each marginal  $X_t$  ( $t \in [0, 1]$ ) admits a representation

$$\mathbb{E}e^{iuX_t} = e^{-t\psi(u)}, \tag{2}$$

where

$$\psi(u) = \frac{\sigma^2}{2}u^2 + ibu + \int_{\mathbb{R} \setminus \{0\}} (1 - e^{iux} + \mathbb{1}_{\{|x| \leq 1\}} iux) \nu(dx)$$

for parameters  $\sigma^2 \in [0, \infty)$ ,  $b \in \mathbb{R}$ , and a positive measure  $\nu$  on  $\mathbb{R} \setminus \{0\}$  with

$$\int_{\mathbb{R} \setminus \{0\}} 1 \wedge x^2 \nu(dx) < \infty. \quad (3)$$

On the other hand, for a given triplet  $(\nu, \sigma^2, b)$  there exists a Lévy process  $X$  such that (2) is valid, moreover the distribution of a Lévy process  $X$  is uniquely characterized by the latter triplet. We will call the corresponding process an  $(\nu, \sigma^2, b)$ -Lévy process.

If (2) is true for

$$\psi(u) = \frac{\sigma^2}{2} u^2 + \int_{\mathbb{R} \setminus \{0\}} (1 - e^{iux} + iux) \nu(dx),$$

then we will call  $X$  a  $(\nu, \sigma^2)$ -Lévy martingale. Note that such a representation implies that  $\int |x| \wedge x^2 \nu(dx)$  is finite and that the Lévy process  $X$  is a martingale in the usual sense.

After stating our main results in Section 1.2, we shall list some important examples in Section 1.3. Then Section 2 is devoted to the analysis of a particular coding scheme. The coding strategy of interest will be a measurable function

$$\Theta = \Theta_{\varepsilon, b, m} : \mathbb{D}[0, 1) \rightarrow \mathbb{D}[0, 1)$$

depending on three parameters  $\varepsilon > 0$ ,  $b \in \mathbb{R}$  and  $m > 0$ . The parameter  $\varepsilon$  will be responsible for the quality of the reconstruction, in the sense that lower  $\varepsilon$  correspond to lower approximation errors. The parameters  $b$  and  $m$  have to be adjusted to  $\varepsilon$  and certain quantities relying on the Lévy measure. Namely, the coding scheme presented below works in a weakly optimal way (in the sense of both quantization constraint and entropy constraint coding error) if  $m = m(\varepsilon)$  is the mean number of jumps to be encoded and  $b = b(\varepsilon)$  is a drift compensation term. If the generating triplet of the Lévy process is given, these parameters are explicitly available for computation. If the generating triplet is not known, these values can be estimated from the data.

In Section 3, we derive lower bounds for the above coding problems. Together, these results show that the provided coding scheme is weakly optimal in many cases.

Throughout, we use the following notation for strong and weak asymptotics. For two functions  $f$  and  $g$ ,  $f(x) \sim g(x)$ , as  $x \rightarrow 0$ , means that  $f(x)/g(x) \rightarrow 1$ , as  $x \rightarrow 0$ . On the other hand, we use the notation  $f(x) \lesssim g(x)$ , as  $x \rightarrow 0$ , if  $\lim_{x \rightarrow 0} f(x)/g(x) \leq 1$ . We also write  $g(x) \gtrsim f(x)$  in this case. Furthermore, we write  $f(x) \approx g(x)$ , as  $x \rightarrow 0$ , if  $0 < \liminf_{x \rightarrow 0} f(x)/g(x) \leq \limsup_{x \rightarrow 0} f(x)/g(x) < \infty$ .

## 1.2 Results

The crucial quantities describing the coding complexity of Lévy processes are

$$F_1(\varepsilon) := \varepsilon^{-2} \left( \sigma^2 + \int x^2 \wedge \varepsilon^2 \nu(dx) \right) \quad \text{and} \quad F_2(\varepsilon) := \int_{[-\varepsilon, \varepsilon]^c} \log(|x|/\varepsilon) \nu(dx).$$

Furthermore, we shall use  $F(\varepsilon) := F_1(\varepsilon) + F_2(\varepsilon)$ . The function integrated by the Lévy measure is visualised in Figure 1. Note that (3) does not ensure the finiteness of  $F_2$  and that  $F_2$  is either finite or infinite for all  $\varepsilon > 0$ .

We are now in a position to state the main results of the article. Let us start with the entropy coding error.

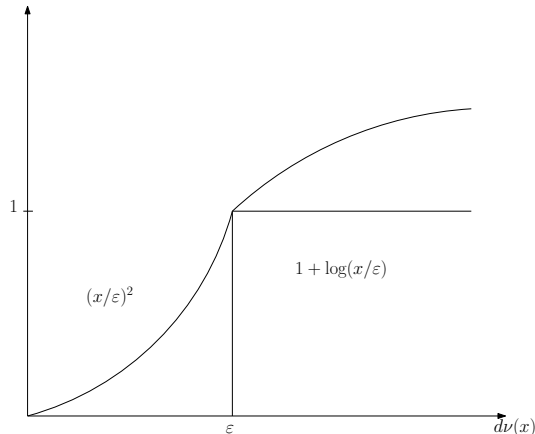


Figure 1: Visualization of the function  $F$

**Theorem 1.1.** *There exist constants  $c_1 = c_1(p) > 0$  and  $c_2 > 0$  such that, for arbitrary Lévy processes with finite  $F_2$ , any  $s > 0$ , and all  $\varepsilon > 0$ ,*

$$D^{(e)}(c_1 F(\varepsilon), s) \leq c_2 \varepsilon.$$

Similarly to the entropy coding error, we obtain the upper bound for the quantization error.

**Theorem 1.2.** *Assume that there is a  $q > s$  such that*

(a)  $\mathbb{E} \|X\|^q < \infty$ ,

(b) for some  $\mu > 0$ ,

$$\limsup_{\varepsilon \rightarrow 0} \frac{\int_{|x| > \varepsilon} (|x|/\varepsilon)^\mu \nu(dx)}{\nu([- \varepsilon, \varepsilon]^c)} < \infty. \quad (4)$$

*Then there exist a constant  $c_1 = c_1(p, \nu) > 0$  and a universal constant  $c_2 > 0$  such that, for all  $0 < \varepsilon < \varepsilon_0 = \varepsilon_0(\nu, s, p)$ ,*

$$D^{(q)}(c_1 F(\varepsilon), s) \leq c_2 \varepsilon.$$

In the proofs of the upper bounds we only need to consider the case where  $F_2$ . Indeed, in the second theorem, assumption (a) implies the finiteness of  $F_2$ .

**Remark 1.3.** Let us comment on the conditions in Theorem 1.2: Condition (a) is natural, though one could soften it by the use of Orlicz norms. Moreover, condition (b) is needed to guarantee that typical realizations of the Lévy process dominate the quantization complexity of the process (see equation (11)). Essentially, (b) does not hold if the Lévy measure is finite or if  $\nu([- \varepsilon, \varepsilon]^c)$  does not grow to infinity fast enough, when  $\varepsilon$  tends to zero.

With given Lévy measure, it is usually easy to verify conditions (a) and (b), cf. Remarks 2.1 and 2.2 below.

**Remark 1.4.** Another approach for the quantization of Lévy processes is taken in [7]. There, linear quantizers are constructed, and a relation of quantization to the path regularity of processes is outlined. However, as observed by Creutzig [2], linear approximations are not optimal whenever  $s > p$ . In this article we work with non-linear quantizers, which lead to better -mostly weakly optimal- results.

The corresponding lower bound reads as follows.

**Theorem 1.5** (Lower bound). *There exist universal constants  $c_1, c_2, c_3 > 0$  such that the following holds. For every Lévy process  $X$  with finite  $F_2$ , any  $\varepsilon > 0$  with  $F_1(\varepsilon) \geq c_3$  one has*

$$D(c_1 F(\varepsilon), 1) \geq c_2 \varepsilon.$$

Moreover, if  $\nu(\mathbb{R}) = \infty$  or  $\sigma \neq 0$ , one has for any  $s > 0$ ,

$$D(c_1 F_1(\varepsilon), s) \gtrsim c_2 \varepsilon$$

as  $\varepsilon \downarrow 0$ . In the case where  $F_2 \equiv \infty$ , one has  $D(r, 1) = \infty$  for any  $r \geq 0$ .

**Remark 1.6.** So far one cannot replace  $F_1$  by  $F$  in the second statement of Theorem 1.5. Since mostly  $F_1$  and  $F$  are weakly equivalent when  $\varepsilon$  tends to zero, the second estimate typically leads to sharp results. Nevertheless, it would be interesting to find out, whether one can close this remaining gap.

Note that we have not specified the basis of the logarithm. However, all results stated above are valid for any basis. The choice of the basis has only an influence on the constants in the theorems. We will work with the basis 2 when proving the upper bounds, since this seems more appropriate in the context of binary representations. When proving the lower bounds we switch to the natural logarithm.

### 1.3 Examples

In this subsection, we apply the above results to some common Lévy processes.

**Example 1.7** (Stable Lévy process). Let us consider the case of an  $\alpha$ -stable Lévy process. Here we have  $\nu(dx) = (C_1 \mathbb{1}_{\{x < 0\}} + C_2 \mathbb{1}_{\{x > 0\}})|x|^{-\alpha-1} dx$ , and one can easily verify that  $F_1(\varepsilon) = C_\alpha^1 \varepsilon^{-\alpha}$  and  $F_2(\varepsilon) = C_\alpha^2 \varepsilon^{-\alpha}$ . All assumptions of the main theorems are satisfied and we conclude that for all moments  $s_1 > 0$ ,  $s_2 \in (0, \alpha)$  and all  $p \geq 1$ ,

$$D(r, s_1) \approx D^{(e)}(r, s_1) \approx D^{(q)}(r, s_2) \approx r^{-1/\alpha}.$$

This improves results from [2] and [7].

Note that the coding complexity  $\alpha$ -stable Lévy process is smaller than the one of a 2-stable Lévy process, i.e. Brownian motion. In fact, this is true for all Lévy process.

**Example 1.8** (Lévy process with non-vanishing Gaussian component). It is easy to calculate that  $F_i(\varepsilon) \leq c\varepsilon^{-2}$  for  $i = 1, 2$ . Therefore, if  $\sigma \neq 0$  then  $F(\varepsilon) \approx F_1(\varepsilon) \approx \varepsilon^{-2}$ .

This has two implications. Firstly, in presence of a Gaussian component, the coding complexity of the Lévy process is the same as for Brownian motion, as long as our results apply. In case  $\sigma = 0$ , the coding complexity is weakly bounded from above by that of Brownian motion.

More precisely,

$$D^{(e)}(r, s) \leq Cr^{-1/2}, \quad \text{for any Lévy process,}$$

and

$$D^{(e)}(r, s) \approx r^{-1/2}, \quad \text{if } \sigma \neq 0.$$

On the other hand, under the assumptions (a) and (b),

$$D^{(q)}(r, s) \leq Cr^{-1/2}, \quad \text{for any Lévy process,}$$

and,

$$D^{(q)}(r, s) \approx r^{-1/2}, \quad \text{if } \sigma \neq 0.$$

In fact, by a modification of (11) one can show that (b) is not necessary if  $\sigma \neq 0$ .

**Example 1.9** (Gamma process). Let us consider a standard Gamma process. In this case,  $\nu(dx) = \mathbb{1}_{\{x>0\}}x^{-1}e^{-x}dx$  and one gets  $F_1(\varepsilon) \approx \log 1/\varepsilon$  and  $F_2(\varepsilon) \approx (\log 1/\varepsilon)^2$ . Consequently, for fixed  $p, s \in [1, \infty)$ , there exist constants  $c_1, c_2, c'_1, c'_2 \in \mathbb{R}_+$  such that for all  $\varepsilon \geq 0$

$$D^{(e)}(c_1(\log 1/\varepsilon)^2, s) \leq c_2\varepsilon$$

and

$$D(c'_1(\log 1/\varepsilon)^2, s) \geq c'_2\varepsilon.$$

Therefore,

$$D(r, s) = \exp(-e^{\mathcal{O}(1)}\sqrt{r}) \quad \text{and} \quad D^{(e)}(r, s) = \exp(-e^{\mathcal{O}(1)}\sqrt{r}).$$

Note that Theorem 1.2 does not apply since condition (4) fails to hold.

**Example 1.10** (Compound Poisson process). Let  $(N(t))_{t \geq 0}$  be a standard Poisson process. Let furthermore  $Y, Y_1, Y_2, \dots$  be i.i.d. random variables that are not a.s. equal to 0 and independent of the Poisson process. Then

$$X(t) := \sum_{i=1}^{N(t)} Y_i$$

is a compound Poisson process, i.e. a Lévy process with Lévy measure  $\nu = \mathbb{P}_Y$  and drift  $b = \mathbb{E}[Y \mathbb{1}_{\{|Y| \leq 1\}}]$ .

It is immediately clear that  $F_1(\varepsilon) \leq 1$  and  $F_2(\varepsilon) \approx \mathbb{E} \left[ \log \left( \frac{|Y|}{\varepsilon} \right) \mathbb{1}_{\{|Y| \geq \varepsilon\}} \right]$  so that  $F_2$  dominates  $F$  when  $\varepsilon$  is small. Thus the main complexity is induced by the “large jumps”. For fixed  $p, s \in [1, \infty)$ , the main theorems imply the existence of constants  $c_1, c_2, c'_1, c'_2 \in \mathbb{R}_+$  such that

$$D^{(e)} \left( c_1 \mathbb{E} \left[ \log \left( \frac{|Y|}{\varepsilon} \right) \mathbb{1}_{\{|Y| \geq \varepsilon\}} \right], s \right) \leq c_2\varepsilon$$

and

$$D \left( c'_1 \mathbb{E} \left[ \log \left( \frac{|Y|}{\varepsilon} \right) \mathbb{1}_{\{|Y| \geq \varepsilon\}} \right], s \right) \geq c'_2\varepsilon$$

Hence,

$$D(r, s) = \exp(-e^{\mathcal{O}(1)}r) \quad \text{and} \quad D^{(e)}(r, s) = \exp(-e^{\mathcal{O}(1)}r).$$

A more precise result for a subclass of compound Poisson processes was already obtained in the dissertation of Vormoor [9]. In particular, in those cases, the rates of quantization and entropy coding error differ.

Note that in the case of a compound Poisson processes we cannot use Theorem 1.2 on the quantization error, since condition (b) is not satisfied.

## 2 Upper Bounds

### 2.1 An Explicit Coding Strategy

In this subsection, we describe an explicit coding strategy that can be used to encode a Lévy process. We derive that the strategy has a mean error of order  $\varepsilon$  and that the bit complexity is given by the quantity in (10). In the following subsections we use this strategy in order to prove upper bounds for the entropy coding error and the quantization error.

The reconstruction  $\hat{X} = \Theta_{\varepsilon, b, m}(X)$  will be a step function with the step heights being integer multiples of  $\varepsilon$ , i.e. we use an  $\varepsilon\mathbb{Z}$  grid to approximate  $X$ . For this purpose, let us define  $g$  to be a nearest neighbour projection of  $\mathbb{R}$  onto  $\varepsilon\mathbb{Z}$ . As a first step, we subtract the drift of the process by setting  $X'(t) := X(t) - b(\varepsilon)t$ , where  $b(\varepsilon)$  is a drift compensation term given by

$$b(\varepsilon) := b - \int_{[-1,1] \setminus [-\varepsilon, \varepsilon]} x \nu(dx) + \int_{[-\varepsilon, \varepsilon] \setminus [-1,1]} x \nu(dx).$$

**Notation.** Set  $S_0 := 0$  and let

$$S_i := \inf \left\{ t > S_{i-1} : |X'_t - g(X'_{S_{i-1}})| \geq 2\varepsilon \right\}; \quad i \in \mathbb{N},$$

be the first exit time of the process  $\left( X'_s - g(X'_{S_{i-1}}) \right)_{s \geq S_{i-1}}$  from the interval  $[-2\varepsilon, 2\varepsilon]$ .

Let  $M := \max\{i : S_i < 1\}$ . Some of the stopping times  $S_i$  are induced by jumps larger than  $\varepsilon$ . These shall be called *large jumps*.

**Coding procedure.** Note that it is possible to detect the jump points  $(S_i)_{i=1, \dots, M}$  by a single swipe through the interval  $[0, 1]$ . For each jump we encode its height and its time separately by using prefix-free representations: we use a prefix-free representation for the integers  $\Upsilon_1 : \mathbb{Z} \setminus \{0\} \rightarrow \{0, 1\}^*$  (as outlined in Lemma 2.4) to code the number  $H_i/\varepsilon \in \mathbb{Z}$ , where  $H_i := g(X'_{S_i}) - g(X'_{S_{i-1}})$  denotes the discretised height. Moreover, the time approximation  $\hat{S}_i$  to  $S_i$  is chosen in such a way that

$$S_i \leq \hat{S}_i < S_{i+1} \quad \text{and} \quad \hat{S}_i - S_i \leq \varepsilon^p / (|H_i|^p M). \quad (5)$$

For a visualization, cf. Figure 2. Concretely, we choose  $\hat{S}_i$  as follows. By Lemma 2.5, there is a coding scheme  $\Upsilon_2 : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \{0, 1\}^*$ , where, for  $r \in [0, 1]$ ,  $\delta > 0$ ,  $\Upsilon_2(r, \delta)$  is the binary representation of a number  $\bar{\Upsilon}_2(r, \delta) \in \bigcup_{n \geq 0} 2^{-n}\mathbb{Z} \cap [0, 1]$  such that  $\bar{\Upsilon}_2(r, \delta) \in [r, r + \delta]$ .

We transmit the information in the following way: we divide the interval  $[0, 1)$  into  $\lceil F_1(\varepsilon) \rceil$  ‘boxes’ (i.e. intervals)  $I_j = [jF_1(\varepsilon), (j+1)F_1(\varepsilon) \wedge 1)$ ,  $j = 0, \dots, \lceil F_1(\varepsilon) \rceil - 1$ . Each jump  $S_i$  ( $i = 1, \dots, M$ ) is translated into the code

$$\pi_i := '0' * \Upsilon_1(H_i/\varepsilon) * \Upsilon_2(F_1(\varepsilon)S_i - \lfloor F_1(\varepsilon)S_i \rfloor, F_1(\varepsilon)(S_{i+1} - S_i) \wedge F_1(\varepsilon)\varepsilon^p / (|H_i|^p M)),$$

where  $*$  denotes the concatenation of strings. Note that  $F_1(\varepsilon)S_i - \lfloor F_1(\varepsilon)S_i \rfloor$  is exactly the difference between the actual jump point and the left corner of the box, scaled on the unit interval. Then each block  $j$  is described by the string

$$\Pi_j := \prod_{\{i: S_i \in I_j\}} \pi_i,$$

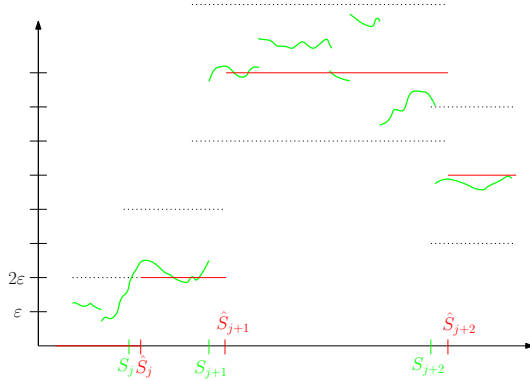


Figure 2: The coding procedure

and finally the complete information is encoded as

$$\prod_{j=0}^{\lceil F_1(\varepsilon) \rceil - 1} (\Pi_j * \text{'1'}).$$

It is easy to check that this provides indeed a prefix-free representation of  $((\hat{S}_i, H_i)_{i=1, \dots, M}, M)$ , and the corresponding approximation defines a deterministic map  $\Theta_{\varepsilon, b(\varepsilon), F_1(\varepsilon)}$  by

$$\hat{X}_t := \Theta_{\varepsilon, b(\varepsilon), F_1(\varepsilon)}(X)(t) := b(\varepsilon)t + \sum_{i=1}^M H_i \mathbb{1}_{\{\hat{S}_i \leq t\}},$$

where

$$\hat{S}_i := a_{\text{box}, i} + \bar{\Upsilon}_2 (F_1(\varepsilon)S_i - \lfloor F_1(\varepsilon)S_i \rfloor, F_1(\varepsilon)(S_{i+1} - S_i) \wedge F_1(\varepsilon)\varepsilon^p / (|H_i|^p M))$$

and  $a_{\text{box}, i}$  is the left corner of the box that contains  $S_i$ . Note that, in order to decode this value, it is sufficient to transmit a code for  $F_1(\varepsilon)S_i - \lfloor F_1(\varepsilon)S_i \rfloor$ . The chosen precision ensures (5). Note that the parameters  $\varepsilon$ ,  $b(\varepsilon)$  and  $F_1(\varepsilon)$  describe the approximation scheme uniquely.

For convenience we will also consider the drift adjusted reconstruction  $\hat{X}'$  defined by

$$\hat{X}'_t := \sum_{i=1}^M H_i \mathbb{1}_{\{\hat{S}_i \leq t\}}.$$

**Waiting time for the jumps.** Let us estimate the waiting time for subsequent jumps. For this purpose, let  $X^{(1)}$  be the process consisting of the (finitely many) jumps of  $X'$  that are greater than  $\varepsilon$  and set  $X^{(2)} := X' - X^{(1)}$ . Note that  $X^{(2)}$  is a  $(\nu_{[-\varepsilon, \varepsilon]}, \sigma^2)$ -Lévy martingale. Denote by  $\Gamma_1$  the stopping time induced by the first jump of  $X^{(1)}$ . Note that  $|X'_{S_{i-1}} - g(X'_{S_{i-1}})| \leq \varepsilon/2$  a.s.



so that due to the strong Markov property one has for all  $t \geq 0$ ,

$$\begin{aligned} \mathbb{P}(S_i - S_{i-1} \leq t \mid \mathcal{F}_{S_{i-1}}) &\leq \mathbb{P}\left(\sup_{0 < s \leq t} |X_s^{(2)}| > \frac{3}{2}\varepsilon\right) + \mathbb{P}(\Gamma_1 \leq t) \\ &\leq (3\varepsilon/2)^{-2} \mathbb{E} \sup_{0 < s \leq t} |X_s^{(2)}|^2 + \nu([- \varepsilon, \varepsilon]^c)t \\ &\leq \varepsilon^{-2} \mathbb{E}|X_t^{(2)}|^2 + \nu([- \varepsilon, \varepsilon]^c)t, \end{aligned}$$

where the last step is justified by Doob's martingale inequality. By the compensation formula ([1], p. 7) the last term equals  $F_1(\varepsilon)t$ .

Let  $U_1, U_2, \dots$  be a sequence of i.i.d. random variables. Then we have shown that for all jumps  $S_i$ ,

$$\mathbb{P}(S_i - S_{i-1} \leq t \mid \mathcal{F}_{S_{i-1}}) \leq \max(tF_1(\varepsilon), 1) = \mathbb{P}(U_i \leq F_1(\varepsilon)t),$$

for all  $t \geq 0$  and  $i \in \mathbb{N}$ . Consequently, we can couple the random times  $(S_i - S_{i-1})_{i \geq 1}$  with the sequence  $(U_i)_{i \geq 1}$  such that

$$F_1(\varepsilon)(S_i - S_{i-1}) \geq U_i. \quad (6)$$

**Coding error.** First, let us analyse the error of the approximation. With  $\tilde{X}' = (g(X'_t))_{t \in [0,1]}$  one gets

$$\|X - \hat{X}\| = \|X' - \hat{X}'\| \leq \underbrace{\|X' - \tilde{X}'\|}_{\leq 2\varepsilon} + \|\tilde{X}' - \hat{X}'\|. \quad (7)$$

Moreover, due to property (5)

$$\|\tilde{X}' - \hat{X}'\|^p = \sum_{i=1}^M |H_i|^p (\hat{S}_i - S_i) \leq \varepsilon^p, \quad (8)$$

so that  $\|X - \hat{X}\| \leq 3\varepsilon$ .

**Coding complexity.** Let us count the number of bits needed in the approximation:

- Each change in a block is indicated by a '1' which gives in total  $\lceil F_1(\varepsilon) \rceil$  bits.
- Each pair  $(H_i, \hat{S}_i)$  is initialized by a '0' which gives in total  $M$  bits.
- Coding the numbers  $H_1/\varepsilon, \dots, H_M/\varepsilon$  by using an appropriate representation  $\Upsilon_1$  needs less than

$$\sum_{i=1}^M 2 \left( 2 + \log \frac{|H_i|}{\varepsilon} \right)$$

bits by Lemma 2.4.

- Coding the numbers  $\hat{S}_1, \dots, \hat{S}_M$  needs less than

$$\sum_{i=1}^M 2 \left( 2 + \log_+ \frac{F_1(\varepsilon)^{-1}}{\varepsilon^p / (M|H_i|^p) \wedge (S_{i+1} - S_i)} \right)$$

bits (see Lemma 2.5).

Therefore, the total bit-length is bounded from above by

$$2 \sum_{i=1}^M \left[ \log \frac{|H_i|}{\varepsilon} + \log_+ \frac{F_1(\varepsilon)^{-1}}{\varepsilon^p / (M|H_i|^p) \wedge (S_{i+1} - S_i)} \right] + 8M + \lceil F_1(\varepsilon) \rceil.$$

This equals

$$2 \sum_{i=1}^M \left[ \log \frac{|H_i|}{\varepsilon} + \log_+ \left( \frac{M|H_i|^p}{F_1(\varepsilon)\varepsilon^p} \vee \frac{1}{F_1(\varepsilon)(S_{i+1} - S_i)} \right) \right] + 8M + \lceil F_1(\varepsilon) \rceil.$$

By (6) and the inequality  $\log_+(x \vee y) \leq \log_+ x + \log_+ y$ , the latter is less than

$$2 \sum_{i=1}^M \left[ (1+p) \log_+ \frac{|H_i|}{\varepsilon} + \log_+ \frac{1}{U_i} \right] + 2M \log_+ \frac{M}{F_1(\varepsilon)} + 8M + \lceil F_1(\varepsilon) \rceil. \quad (9)$$

Next, recall from (6) that  $F_1(\varepsilon)(S_i - S_{i-1}) \geq U_i$  so that

$$F_1(\varepsilon)M \geq \sum_{i=1}^M F_1(\varepsilon)(S_i - S_{i-1}) \geq \sum_{i=1}^M U_i;$$

and using the convexity of  $\log_+(1/\cdot)$  one gets with Jensen's Inequality

$$\sum_{i=1}^M \log \frac{1}{U_i} = M \sum_{i=1}^M \frac{1}{M} \log \frac{1}{U_i} \geq M \log_+ \frac{1}{\sum_{i=1}^M \frac{U_i}{M}} \geq M \log_+ \frac{M}{F_1(\varepsilon)}.$$

We conclude with (9) that

$$2 \sum_{i=1}^M \left[ (1+p) \log_+ \frac{|H_i|}{\varepsilon} + 2 \log \frac{1}{U_i} \right] + 8M + \lceil F_1(\varepsilon) \rceil$$

is an upper bound for the bit-length.

We conclude with (9) that

$$2 \sum_{i=1}^M \left[ (1+p) \log_+ \frac{|H_i|}{\varepsilon} + 2 \log \frac{1}{U_i} \right] + 8M + \lceil F_1(\varepsilon) \rceil$$

is an upper bound for the bit-length. Denoting for any time  $t > 0$  the jump at time  $t$  by  $\Delta X_t = X_t - X_{t-}$  allows us to estimate  $|H_i| \leq |\Delta X_{S_i}| + \frac{5}{2}\varepsilon$  so that basic analysis gives

$$\log_+ \frac{|H_i|}{\varepsilon} \leq 5 + \log_+ \frac{|\Delta X_{S_i}|}{\varepsilon}.$$

Consequently, the bit-length is bounded by

$$\begin{aligned} & 4 \sum_{i=1}^M \log \frac{1}{U_i} + 2(1+p) \sum_{t \in (0,1]} \log_+ \frac{|\Delta X_t|}{\varepsilon} + (18+10p)M + \lceil F_1(\varepsilon) \rceil \\ & \leq K_1(p) \sum_{i=1}^M \left[ 1 + \log \frac{1}{U_i} \right] + K_2(p) \sum_{t \in (0,1]} \log_+ \frac{|\Delta X_t|}{\varepsilon} + F_1(\varepsilon) + 1, \end{aligned} \quad (10)$$

where  $K_1(p)$  and  $K_2(p)$  are constants only depending on  $p$ .

## 2.2 Proof of Theorem 1.1

**Proof.** By (7) and (8) the error (and thus the mean error, for all moments  $s > 0$ ) is less than  $3\varepsilon$ .

On the other hand, the coding complexity of the algorithm constructed above is given by (10). Let us look at what the different terms amount to on average. Note that

$$\mathbb{E} \sum_{t \in (0,1]} \log_+ \frac{|\Delta X_t|}{\varepsilon} = F_2(\varepsilon),$$

by the compensation formula ([8], p. 29). Finally, by Lemma 2.3, we have

$$\mathbb{E} \sum_{i=1}^M (1 + \log U_i^{-1}) \leq cF_1(\varepsilon).$$

This shows that the expected bit length of the whole message is less than  $c_1 F(\varepsilon)$ , with some constant  $c_1$  depending only on  $p$ , as required.  $\square$

## 2.3 Proof of Theorem 1.2

**Proof.** We use the coding scheme explained above. However, we encode by the zero function in case that the number of small jumps,  $M$ , exceeds  $C_1 F(\varepsilon)$ , where  $C_1$  is a constant to be chosen presently. The same is done if the complexity to encode the jump heights of the large jumps, namely  $\sum_{t \in (0,1]} \log_+ |\Delta X_t|/\varepsilon$ , or the complexity to encode the positions of the jumps, namely  $\sum_{i=1}^M (1 + \log U_i^{-1})$ , is larger than  $C_2 F(\varepsilon)$ , where  $C_2$  is a constant to be chosen presently. Let us define  $\mathcal{T}$  to be the event that none of the above cases occurs, i.e. the ‘typical case’.

Note that, by the exponential compensation formula ([1], p. 8),

$$\begin{aligned} \mathbb{P} \left( \sum_{t \in (0,1]} \log_+ \frac{|\Delta X_t|}{\varepsilon} > C_2 F(\varepsilon) \right) &\leq e^{-C_2 \mu F(\varepsilon)} \mathbb{E} e^{\mu \sum_{t \in (0,1]} \log_+ \frac{|\Delta X_{S_t}|}{\varepsilon}} \\ &\leq e^{-C_2 \mu F(\varepsilon)} e^{-\int_{|x| \geq \varepsilon} 1 - (|x|/\varepsilon)^\mu \nu(dx)} \leq e^{-C_2 \mu F(\varepsilon)} e^{EF(\varepsilon)} \leq e^{-C_2/2 \mu F(\varepsilon)}, \end{aligned} \quad (11)$$

where  $E$  is some constant depending on the finite constant in (4) only. The last step holds for  $C_2$  large enough. On the other hand, by the Chebyshev Inequality,

$$\mathbb{P} \left( \sum_{i=1}^{C_1 F(\varepsilon)} (1 + \log U_i^{-1}) > C_2 F(\varepsilon) \right) \leq e^{-C_2/2 F(\varepsilon)},$$

for  $C_2$  large enough. Finally, one proves, e.g. using the same discretization as in (13), that for  $C_1$  large enough,

$$\mathbb{P}(M > C_1 F(\varepsilon)) \leq e^{-C_1/2 F(\varepsilon)}.$$

Therefore, for some positive constant  $C$  depending on  $\mu$  and  $E$ , we have  $\mathbb{P}(\mathcal{T}^c) \leq \exp(-CF(\varepsilon))$ .

Let  $r > 0$  be chosen by  $1/q + 1/r = 1/s$ . Let  $\kappa > 0$  be chosen small enough such that  $C\nu([- \kappa, \kappa]^c) \geq r$ . This is possible, since  $\nu([- \kappa, \kappa]^c)$  tends to infinity when  $\kappa \rightarrow 0$ , by condition (b). Then, for  $\varepsilon < \kappa$ ,

$$F(\varepsilon) \geq F_2(\varepsilon) = \int_{[-\varepsilon, \varepsilon]^c} \log \frac{|x|}{\varepsilon} \nu(dx) \geq \nu([- \kappa, \kappa]^c) \log \frac{1}{\varepsilon} \geq -\frac{1}{C} \log \varepsilon^r.$$

Thus,

$$\mathbb{P}(\mathcal{T}^c) \leq e^{-CF(\varepsilon)} \leq \varepsilon^r. \quad (12)$$

Note that the bit complexity of our algorithm is constant if  $\mathcal{T}^c$  occurs and, by (10), less than  $CF(\varepsilon)$  if  $\mathcal{T}$  occurs, where  $C$  depends on  $\mu$  and  $E$ . Then we have for the mean error, using the Hölder Inequality and  $s \geq 1$ ,

$$\begin{aligned} \left( \mathbb{E} \|X - \hat{X}\|^s \right)^{1/s} &\leq \left( \mathbb{E} \mathbf{1}_{\mathcal{T}} \|X - \hat{X}\|^s \right)^{1/s} + \left( \mathbb{E} \mathbf{1}_{\mathcal{T}^c} \|X - \hat{X}\|^s \right)^{1/s} \\ &\leq c_2 \varepsilon + (\mathbb{E} \mathbf{1}_{\mathcal{T}^c}^r)^{1/r} (\mathbb{E} \|X\|^q)^{1/q} \\ &\leq c_2 \varepsilon \left[ 1 + \varepsilon^{-1} c_2^{-1} \mathbb{P}(\mathcal{T}^c)^{1/r} (\mathbb{E} \|X\|^q)^{1/q} \right], \end{aligned}$$

where the term in brackets is bounded, by assumption (a) and (12). Note that the argument works analogously for  $0 < s < 1$ .  $\square$

**Remark 2.1.** It is easy to see that condition (a) is equivalent to the condition

$$\int_{|x|>1} |x|^q \nu(dx) < \infty.$$

**Remark 2.2.** Let us assume that (a) holds. A sufficient condition for (b) to hold is that  $\nu([-2\varepsilon, 2\varepsilon]^c) \leq c \cdot \nu([- \varepsilon, \varepsilon]^c)$  for some  $0 < c < 1$  and all  $0 < \varepsilon \leq \varepsilon_0$ . This can be seen as follows:

$$\begin{aligned} \int_{\varepsilon < |x| \leq \varepsilon_0} \left( \frac{|x|}{\varepsilon} \right)^\mu \nu(dx) &\leq \sum_{k=0}^{\log(\varepsilon_0/\varepsilon)} \int_{2^k \varepsilon < |x| \leq 2^{k+1} \varepsilon} \left( \frac{|x|}{\varepsilon} \right)^\mu \nu(dx) \\ &\leq \sum_{k=0}^{\log(\varepsilon_0/\varepsilon)} \nu([-2^k \varepsilon, 2^k \varepsilon]^c) 2^{(k+1)\mu} \leq \sum_{k=0}^{\infty} c^k 2^{(k+1)\mu} \nu([- \varepsilon, \varepsilon]^c). \end{aligned}$$

Choosing  $0 < \mu < (-\log c) \wedge q$  yields

$$\int_{|x|>\varepsilon} \left( \frac{|x|}{\varepsilon} \right)^\mu \nu(dx) \leq K(\mu, c) \nu([- \varepsilon, \varepsilon]^c) + \varepsilon^{-\mu} \int_{\varepsilon_0 < |x| \leq 1} |x|^\mu \nu(dx) + \varepsilon^{-\mu} \int_{|x|>1} |x|^q \nu(dx),$$

which implies (4).

Note that, in particular, this is the case if  $\varepsilon \mapsto \nu([- \varepsilon, \varepsilon]^c)$  is regularly varying at zero with negative exponent.

## 2.4 Technical tools

In this section, we prove some technical tools that are needed in the proofs of the main results.

**Lemma 2.3.** *Let  $\lambda > 0$  and let  $(U_i)_{i \geq 1}$  be an i.i.d. sequence of random variables uniformly distributed in  $[0, 1]$ . For  $N := \min\{n \in \mathbb{N} : \sum_{i=1}^n U_i \geq \lambda\}$  one has*

$$\mathbb{E} \sum_{i=1}^N (1 + \log U_i^{-1}) \leq 6 \lceil 2\lambda \rceil.$$

**Proof.** Let  $0 \leq s \leq 1$ . Define  $N(s) := \min\{n \in \mathbb{N}_0 : \sum_{i=1}^n U_i \geq s\}$  and consider the function

$$\Psi(s) := \mathbb{E} \sum_{i=1}^{N(s)} (1 + \log U_i^{-1}).$$

We are interested in  $\Psi(\lambda)$ . Clearly,  $\Psi(s) = 0$  for  $s \leq 0$  and  $\Psi$  is increasing. Moreover, one has for  $s > 0$ ,

$$\begin{aligned} \Psi(s) &= \int \left( 1 + \log x^{-1} + \mathbb{E} \sum_{i=1}^{N(s-x)} (1 + \log U_i^{-1}) \right) d\mathbb{P}_{U_1}(x) \\ &= 1 + \int_0^1 -\log x \, dx + \int \mathbb{E} \sum_{i=1}^{N(s-x)} (1 + \log U_i^{-1}) d\mathbb{P}_{U_1}(x) \\ &= 1 + \log e + \int \Psi(s-x) d\mathbb{P}_{U_1}(x) \leq 3 + \int \Psi(s-x) d\mathbb{P}_{U_1}(x). \end{aligned}$$

Let us define

$$U'_1 := \begin{cases} 0 & U_1 \leq 1/2 \\ 1/2 & U_1 > 1/2. \end{cases} \quad (13)$$

Then  $U'_1 \leq U_1$ ; and since  $\Psi$  is increasing, we have

$$\Psi(s) \leq 3 + \int \Psi(s-x) \mathbb{P}_{U'_1}(x) = 3 + \frac{1}{2} \Psi(s) + \frac{1}{2} \Psi\left(s - \frac{1}{2}\right).$$

Therefore,  $\Psi(s) \leq 6 + \Psi\left(s - \frac{1}{2}\right)$  and we get that

$$\Psi(\lambda) \leq 6 + \Psi\left(\lambda - \frac{1}{2}\right) \leq 6 + 6 + \Psi(\lambda - 1) \leq \dots \leq 6 \cdot \lceil 2\lambda \rceil.$$

□

Let us finally gather two facts concerning the coding of integers and real numbers from a given interval, respectively.

**Lemma 2.4.** *There is a universal coding scheme that returns a prefix free code  $\Upsilon_1(x) \in \{0, 1\}^*$  for a given integer  $x \in \mathbb{Z}$  that has a length of at most  $2(2 + \log x)$  bits.*

**Proof.** The sign is encoded by a first bit. Thus, assume  $x > 0$ , because  $x = 0$  can be encoded by ‘00’. Let  $n := \min\{l \in \mathbb{N} \mid x < 2^l\}$ . Then  $2^{n-1} \leq x < 2^n$ . Consider the representation of  $x$  in the binary system. Because of the definition of  $n$ , this representation must have  $n$  bits, the first one of which is a ‘1’.

A prefix free code for  $x$  is given by  $n$  times ‘1’, followed by a ‘0’ and the  $n - 1$  bit long representation of  $x$  in the binary system having taken away the redundant leading ‘1’.

The length of the code is  $2n + 1$ , which is less than  $2(1 + \log_+ x)$ .  $\square$

Let us remark that Lemma 2.4 can be improved up to the order  $\log x + C \log \log x + D$ , as shown in [4].

**Lemma 2.5.** *There exists a universal coding strategy  $\Upsilon_2 : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \{0, 1\}^*$  such that, for any  $\delta > 0$  and  $r \in [0, 1]$ ,  $\Upsilon_2$  returns the prefix free binary representation  $\Upsilon_2(r, \delta)$  of a number  $\overline{\Upsilon}_2(r, \delta) \in [r, r + 1]$  with  $r \leq \overline{\Upsilon}_2(r, \delta) \leq r + \delta$  that needs at most  $2(2 - \log \delta)$  bits.*

**Proof.** Let  $N := \min\{n : \delta \geq 2^{-n-1}\}$ . We choose  $\overline{\Upsilon}_2(1, r, \delta) \in [r, r + 1] \cap S_N$  nearest possible, but larger than  $r$ , where

$$S_N := \bigcup_{n=0}^N 2^{-n}\mathbb{Z}.$$

This ensures that  $0 \leq \overline{\Upsilon}_2(1, r, \delta) - r \leq 2^{-(N+1)} \leq \delta$ , as required.

Any number  $\hat{r} \in [0, 1] \cap S_N$  has a unique representation  $\hat{r} = k2^{-n}$ , with  $k$  uneven,  $1 \leq k \leq 2^n - 1$ ,  $1 \leq n \leq N$ . As a prefix free code  $\Upsilon_2(1, r, \delta)$  for  $\overline{\Upsilon}_2(1, r, \delta)$  we chose the prefix free code for the integer  $2^{n-1} + (k + 1)/2$ . Since  $\overline{\Upsilon}_2(\gamma, \delta) \in S_N$ , we have to encode integers from 2 up to at most  $2^N$ , which, by Lemma 2.4, requires at most  $2(1 + N)$  bits, which is less than  $2(2 - \log \delta)$  bits, by the definition of  $N$ .  $\square$

### 3 Lower bound

The aim of this section is to provide lower bounds for the distortion rate function of the Lévy process. The analysis is divided into three subsections. First we introduce some concepts of information theory and we prove some preliminary results. Next, we provide a lower bound based on  $F_2$ . In the last subsection we give a lower bound in terms of  $F_1$ . Both lower bounds then immediately imply Theorem 1.5.

So far  $p$  is a fixed value in  $[1, \infty)$ . Since the distortion rate function is increasing in the parameter  $p$ , we can and will fix  $p = 1$  in the following discussion.

As mentioned before, we can freely choose the basis of the logarithm in the proof of the main theorems. For the rest of this article, we fix as basis  $e$ .

#### 3.1 Preliminaries

First we will introduce some concepts of information theory. We will need the concept of conditional mutual information. Let  $A, B$  and  $C$  denote random vectors attaining values in some Borel spaces. Then one defines the mutual information between  $A$  and  $B$  given  $C$  as

$$I(A; B|C) = \int I(A; B|C = c) d\mathbb{P}_C(c),$$

where

$$I(A; B|C = c) = \begin{cases} \int \log \frac{d\mathbb{P}_{A,B|C=c}}{d\mathbb{P}_{A|C=c} \otimes d\mathbb{P}_{B|C=c}} d\mathbb{P}_{A,B|C=c} & \text{if } \mathbb{P}_{A,B|C=c} \ll \mathbb{P}_{A|C=c} \otimes \mathbb{P}_{B|C=c} \\ \infty & \text{otherwise.} \end{cases}$$

A summary of computation rules for the mutual information can be found in [5].

**Lemma 3.1.** *For  $n \in \mathbb{N}$ , let  $Y_0, \dots, Y_{n-1}$  and  $\hat{Y}_0, \dots, \hat{Y}_{n-1}$  and  $H$  denote random variables in possibly different Borel spaces. We write shortly  $Y = (Y_0, \dots, Y_{n-1})$ ,  $Y^i = (Y_0, \dots, Y_i)$  for  $0 \leq i \leq n-1$  and  $\hat{Y} = (\hat{Y}_0, \dots, \hat{Y}_{n-1})$ . Then one has*

$$I(Y, H; \hat{Y}) \geq I(Y_0; \hat{Y}_0|H) + I(Y_1; \hat{Y}_1|H, Y^0) + \dots + I(Y_{n-1}; \hat{Y}_{n-1}|H, Y^{n-2}).$$

Moreover, we will need to evaluate the distortion rate function for other originals than the Lévy process  $X$  and for other distortions than  $L_p[0, 1]$ -norm. For a measure  $\mu$  on a Borel space  $E$  and a measurable function  $\rho : E \times E \rightarrow [0, \infty]$  (*distortion measure*) we write

$$D(r|\mu, \rho) = \inf \{ \mathbb{E}[\rho(X, \hat{X})] : \hat{X} \text{ } E\text{-valued r.v. with } I(X; \hat{X}) \leq r \}.$$

Moreover, we associate to a map  $\rho : E \rightarrow [0, \infty]$  the *difference distortion measure*  $\rho : E \times E \rightarrow [0, \infty]$  (denoted by the same identifier) given as  $\rho(x, \hat{x}) = \rho(x - \hat{x})$ . Sometimes we will also consider a general moment  $s > 0$  and write

$$D(r|\mu, \rho, s) = \inf \{ \mathbb{E}[\rho(X, \hat{X})^s]^{1/s} : \hat{X} \text{ } E\text{-valued r.v. with } I(X; \hat{X}) \leq r \}.$$

Moreover, we will omit  $\rho$  if it is the norm based distortion induced by the  $L^1[0, 1]$ -norm.

The following proposition allows us to separately consider the influence of the large jumps and the diffusive part with small jumps onto the coding complexity of the Lévy process:

**Proposition 3.2.** *Let  $E$  be a Borel-space and assume that  $(E, +)$  is an Abelian group such that the sum is Borel-measurable. Denote by  $A$  and  $B$  independent  $E$ -valued random elements and suppose that there exists a measurable map  $\varphi : E \rightarrow E^2$  with*

$$\varphi(A + B) = (A, B) \text{ a.s.} \tag{14}$$

*Then, under any difference distortion measure  $\rho$  on  $E$ , one has for every  $r \geq 0$ :*

$$D(r|\mathbb{P}_{A+B}, \rho) \geq D(r|\mathbb{P}_A, \rho).$$

**Proof.** Fix  $r \geq 0$ . Next, we use that the distortion rate function  $D(\cdot|\mathbb{P}_A, \rho)$  is convex. We denote by  $f$  a tangent of  $D(\cdot|\mathbb{P}_A, \rho)$  at the point  $r$ . Then, for any random element  $Z$  on  $E$ ,

$$\begin{aligned} \mathbb{E}[\rho(A + B, Z)] &= \int \mathbb{E}[\rho(A, Z - b)|B = b] dP_B(b) \\ &\geq \int f(I(A; Z|B = b)) dP_B(b) = f\left(\int I(A; Z|B = b) dP_B(b)\right) \\ &= f(I(A; Z|B)). \end{aligned}$$

Therefore,

$$\inf_{\{Z: I(A; Z|B) \leq r\}} \mathbb{E}[\rho(A + B, Z)] \geq f(r) = D(r|\mathbb{P}_A, \rho).$$

On the other hand, by assumption (14),  $I(A + B; Z) = I((A, B); Z)$  for any random element  $Z$  on  $E$ . Hence,

$$I(A + B; Z) = I((A, B); Z) = I(B; Z) + I(A; Z|B) \geq I(A; Z|B).$$

Therefore,

$$\begin{aligned} D(r|\mathbb{P}_{A+B}, \rho) &= \inf_{\{Z: I(A+B; Z) \leq r\}} \mathbb{E}[\rho(A + B, Z)] \\ &\geq \inf_{\substack{\{Z \text{ r.v. on } E: \\ I(A; Z|B) \leq r\}}} \mathbb{E}[\rho(A + B, Z)] \geq D(r|\mathbb{P}_A, \rho). \end{aligned}$$

□

### 3.2 Lower bound based on $F_2$

**Theorem 3.3.** *There exists some universal constant  $c$  such that for all  $\varepsilon > 0$ ,*

$$D\left(\frac{\kappa(\varepsilon)}{e} F_2(\varepsilon) \middle| X, L_1[0, 1], 1\right) \geq c \kappa(\varepsilon) \varepsilon,$$

where  $\kappa(\varepsilon) = \kappa(\varepsilon, \nu) = \lfloor \nu([- \varepsilon, \varepsilon]^c) \rfloor / \nu([- \varepsilon, \varepsilon]^c)$ .

The proof of the theorem is based on the following idea: in order to find an approximation of accuracy  $\varepsilon$ , one needs to allocate about  $\log_+ |X_t - X_{t-}| / \varepsilon$  bits (nats) for each big jump.

The problem is related to a minimization problem that we want to introduce now. Let  $\Pi$  be a finite non-negative measure on a measurable space  $(E, \mathcal{E})$  and let  $h : E \rightarrow [0, \infty)$  denote a Borel-measurable function with

$$\int \log_+ h(x) d\Pi(x) < \infty.$$

The aim is now to minimize for given  $r > 0$  the target function

$$\int h(x) \exp(-\xi(x)) \Pi(dx)$$

over all measurable functions  $\xi : E \rightarrow [0, \infty)$  satisfying the constraint

$$\int \xi(x) d\Pi(x) \leq r. \tag{15}$$

**Lemma 3.4.** *Assuming that  $\{h > 0\}$  has not  $\Pi$ -measure zero, the minimization problem possesses a  $\Pi$ -a.e. unique solution of the form*

$$\xi(x) = \log_+ \frac{h(x)}{\lambda}, \tag{16}$$

where  $\lambda = \lambda(r) > 0$  is an appropriate parameter depending on  $r > 0$ . When the optimal function  $\xi$  is as in (16), then the minimal value of the target function is

$$\int \lambda \wedge h(x) \Pi(dx).$$



**Proof.** The proof is based on a Lagrangian analysis. Let  $\zeta(y) = \exp(-y)$  ( $y \in [0, \infty)$ ) and consider its convex conjugate

$$\bar{\zeta}(z) = \inf_{y \geq 0} [\zeta(y) + yz] \quad (z \geq 0).$$

Let  $\lambda > 0$  and denote by  $\tilde{\Pi}$  the  $\sigma$ -finite measure with  $\frac{d\tilde{\Pi}}{d\Pi}(x) = h(x)$ . Now observe that for a non-negative function  $\xi$  satisfying the constraint (15) one has

$$\int h(x) \exp(-\xi(x)) d\Pi(x) \geq \int \left[ \zeta(\xi(x)) + \lambda \frac{\xi(x)}{h(x)} \right] d\tilde{\Pi}(x) - \lambda r \quad (17)$$

$$\geq \int \bar{\zeta} \left( \frac{\lambda}{h(x)} \right) d\tilde{\Pi}(x) - \lambda r. \quad (18)$$

The last expression in this estimate does not depend on the choice of  $\xi$ . If we can establish equality in the above estimates for certain  $\xi$  and  $\lambda$ , then this  $\xi$  minimizes the problem.

Next, we note that one has equality in (17) iff

$$\begin{cases} \int \xi(x) d\Pi(x) = r \text{ and} \\ \xi(x) = 0 \text{ for } \Pi\text{-a.e. } x \text{ with } h(x) = 0. \end{cases} \quad (19)$$

We need to look for a non-negative function  $\xi$  and a parameter  $\lambda > 0$  such that (19) is valid and such that

$$\bar{\zeta} \left( \frac{\lambda}{h(x)} \right) = \zeta(\xi(x)) + \frac{\lambda}{h(x)} \xi(x) \text{ for } \tilde{\Pi}\text{-a.e. } x. \quad (20)$$

It is straightforward to verify that for positive  $z$  the function

$$[0, \infty) \ni y \mapsto \zeta(y) + zy \in (0, \infty)$$

attains its unique minimum in  $y = \log_+ \frac{1}{z}$ . Therefore, condition (20) is equivalent to

$$\xi(x) = \xi_\lambda(x) := \log_+ \frac{h(x)}{\lambda} \text{ for } \tilde{\Pi}\text{-a.e. } x.$$

Together with (19) a sufficient criterion for  $\xi$  being a minimum is the existence of a  $\lambda > 0$  such that

$$\begin{cases} \int \xi(x) d\Pi(x) = r \text{ and} \\ \xi(x) = \xi_\lambda(x) \text{ for } \Pi\text{-a.e. } x. \end{cases}$$

Such a  $\lambda$  exists since the function

$$g : (0, \infty) \ni \lambda \mapsto \int \xi_\lambda(x) dx \in [0, \infty)$$

is continuous (due to the dominated convergence theorem) and satisfies

$$\lim_{\lambda \downarrow 0} g(\lambda) = \infty \quad \text{and} \quad \lim_{\lambda \rightarrow \infty} g(\lambda) = 0.$$

Note that if  $\xi$  does not coincide with  $\xi_\lambda$   $\Pi$ -a.e. (where  $\lambda$  is such that  $g(\lambda) = r$ ), then one of the inequalities (17) or (18) is a strict inequality so that  $\xi$  does not minimize the target function.  $\square$

**Proof of Theorem 3.3.** Fix  $\varepsilon > 0$ . Due to Proposition 3.2 we can assume without loss of generality that  $X$  is a pure jump process with jumps bigger than  $\varepsilon$ . Next, let  $l = 1/\nu([-\varepsilon, \varepsilon]^c)$ ,  $n = \lfloor 1/l \rfloor$  and

$$r = \frac{nl}{e} \int_{[-\varepsilon, \varepsilon]^c} \log \frac{|x|}{\varepsilon} \nu(dx) = \frac{\kappa(\varepsilon)}{e} F_2(\varepsilon).$$

We will prove that for an arbitrarily fixed reconstruction  $\hat{X}$  with  $I(X; \hat{X}) \leq r$  one has

$$\mathbb{E}[\|X - \hat{X}\|_{L_1[0,1]}] \geq cnl\varepsilon,$$

where  $c > 0$  is a universal constant.

We let

$$\pi : L_1[0, 1] \rightarrow \ell_1^n, \quad (x_t) \mapsto \left( \left| \int_{il}^{(i+1)l} (2\mathbf{1}_{\{t \geq (2i+1)l/2\}} - 1)x_t \frac{dt}{l} \right| \right)_{i=0, \dots, n-1}$$

and consider

$$Y = (Y_i)_{i=0, \dots, n-1} = \pi(X) \quad \text{and} \quad \hat{Y} = \pi(\hat{X}).$$

The map  $\pi$  is  $l^{-1}$ -Lipschitz continuous so that

$$\mathbb{E}[\|Y - \hat{Y}\|_{\ell_1^n}] \leq l^{-1} \mathbb{E}[\|X - \hat{X}\|]. \quad (21)$$

Moreover,  $\pi$  is invariant under uniform shifts on each time interval  $[i/n, (i+1)/n]$  so that in particular,

$$\pi(X) = \pi\left(X - \sum_{i=0}^{n-1} X_{\frac{2i+1}{2}l} \mathbf{1}_{[il, (i+1)l)}\right).$$

Due to the strong Markov property of the Lévy process, the random variables  $Y_0, \dots, Y_{n-1}$  are i.i.d. We shall derive a lower bound for  $\mathbb{E}[\|Y - \hat{Y}\|_{\ell_1^n}]$ .

For  $i = 0, \dots, n-1$  consider the events

$$A_i = \{X \text{ contains in } [il, (i+1)l) \text{ exactly one jump}\}.$$

and the random vector  $H = (H_i)_{i=0, \dots, n-1}$  given by

$$H_i = \begin{cases} \text{size of the jump in } [il, (i+1)l) & \text{if } A_i \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Next, denote  $Y^i = (Y_0, \dots, Y_i)$  for  $i = 0, \dots, n-1$  and  $Y^{-1} = 0$ . Our objective is to find a lower bound for

$$\mathbb{E}[\|Y - \hat{Y}\|_{\ell_1^n}] \geq \mathbb{E}\left[\sum_{i=0}^{n-1} \mathbb{E}[\|\mathbf{1}_{A_i} Y_i - \mathbf{1}_{A_i} \hat{Y}_i\| | H, Y^{i-1}]\right]. \quad (22)$$

For each  $i \in \{0, \dots, n-1\}$  we analyze the inner expectation. Let  $f_i(h, y^{i-1}) = I(Y_i, \hat{Y}_i | H = h, Y^{i-1} = y^{i-1})$  and consider the random variable

$$R_i = f_i(H, Y^{i-1}).$$

Given  $H$  and  $Y^{i-1}$ , the r.v.  $Y_i$  is uniformly distributed on  $[\mathbb{1}_{\{H_i < 0\}}H_i/2, \mathbb{1}_{\{H_i > 0\}}H_i/2]$ . Therefore,

$$\mathbb{E}[\|\mathbb{1}_{A_i}Y_i - \mathbb{1}_{A_i}\hat{Y}_i \mid |H, Y^{i-1}] \geq D(R_i|\mathcal{U}[0, |H_i|/2], |\cdot|), \quad (23)$$

where  $\mathcal{U}[0, u]$  denotes the uniform distribution on  $[0, u]$ . Now there exists a universal constant  $c > 0$  such that for any  $\bar{r} \geq 0$  and any  $u \geq 0$

$$D(\bar{r}|\mathcal{U}[0, u/2], |\cdot|) \geq cu e^{-\bar{r}}.$$

Together with (22) and (23) we arrive at

$$\mathbb{E}[\|Y - \hat{Y}\|_{\ell_1^n}] \geq c\mathbb{E} \sum_{i=0}^{n-1} |H_i| e^{-R_i}.$$

With  $\Pi$  defined as the product measure  $\mathbb{P} \otimes \sum_{j=0}^{n-1} \delta_j$  we get

$$\mathbb{E}[\|Y - \hat{Y}\|_{\ell_1^n}] \geq c \int |H_i| e^{-R_i} d\Pi(\omega, i). \quad (24)$$

On the other hand, one has  $\mathbb{E}[R_i] = I(Y_i, \hat{Y}_i | H, Y^{i-1})$  by definition so that by Lemma 3.1

$$\int R_i d\Pi(\omega, i) = \sum_{i=0}^{n-1} \mathbb{E}[R_i] \leq I(Y, H; \hat{Y}) \leq I(X; \hat{X}) \leq r.$$

Now consider the minimization problem for the target function

$$\int |H_i| e^{-R_i} d\Pi(\omega, i),$$

where the minimum is taken over all random variables  $R_i$  ( $i = 0, \dots, n-1$ ) satisfying  $\int R_i d\Pi(\omega, i) \leq r$ . The law of  $H_i$  is  $(1 - e^{-1})\delta_0 + \frac{1}{e\nu([-\varepsilon, \varepsilon]^c)}\nu|_{[-\varepsilon, \varepsilon]^c}$  so that

$$\int \log_+ \frac{|H_i|}{\varepsilon} d\Pi(\omega, i) = \frac{n}{e} \int_{[-\varepsilon, \varepsilon]^c} \log \frac{|x|}{\varepsilon} \frac{\nu(dx)}{\nu([-\varepsilon, \varepsilon]^c)} = r.$$

Hence, Lemma 3.4 implies that the optimal value in the minimization problem is

$$\int \varepsilon \mathbb{1}_{\{h_j \neq 0\}} d\Pi(\omega, j) = \frac{1}{e} n \varepsilon.$$

Together with (21) and (24) we get that

$$\mathbb{E}[\|X - \hat{X}\|] \geq \frac{c}{e} \ln \varepsilon$$

which yields the assertion. □

### 3.3 Lower bound related to the $F_1$ -term

**Theorem 3.5.** *There exist positive universal constants  $c_1$  and  $c_2$  such that the following statements are true. For any  $\varepsilon > 0$  with  $F_1(\varepsilon) \geq 18$ , one has*

$$D(c_1 F_1(\varepsilon), 1) \geq c_2 \varepsilon.$$

*If  $\nu(\mathbb{R}) = \infty$  or  $\sigma \neq 0$ , then for any  $s > 0$ , one has*

$$D(c_1 F_1(\varepsilon), s) \gtrsim c_2 \varepsilon$$

*as  $\varepsilon \downarrow 0$ .*

Let us give some heuristics on the proof of the theorem. As we have mentioned before the drift adjusted process  $X'$  needs approximately the time  $1/F_1(\varepsilon)$  to leave an interval of length  $2\varepsilon$ . Assuming that the process is symmetric the process leaves the stripe to either of the sides with equal probability (here one also needs to assume that one starts in the center of the interval). Thus in order to have a coding of accuracy  $\varepsilon$  one needs to describe at least in which direction the process left the stripe for most of the exits. This requires about  $F_1(\varepsilon)$  bits.

As the following remark explains, it suffices to prove the theorem for symmetric Lévy processes.

**Remark 3.6.** Let  $X^*$  denote an independent copy of  $X$  and observe that for  $s \in (0, 1]$

$$D(2r \mid \mathbb{P}_{X-X^*}, s) \leq 2^{1/s} D(r \mid \mathbb{P}_X, s).$$

The process  $X - X^*$  is a symmetric Lévy process and the functions describing its complexity are

$$\tilde{F}_1(\varepsilon) = 2F_1(\varepsilon) \text{ and } \tilde{F}_2(\varepsilon) = 2F_2(\varepsilon).$$

We assume from now on that the Lévy process  $X$  has no drift and a symmetric Lévy measure  $\nu$ .

**Lemma 3.7.** *Let  $\varepsilon > 0$  and denote*

$$T = \inf\{t \geq 0 : |X_t| \geq \varepsilon\}.$$

*Then*

$$\mathbb{P}(T \geq t) \leq \frac{9}{4F_1(2\varepsilon)t}.$$

**Proof.** We consider a Lévy process  $X^*$  with Lévy measure  $\nu^* = \nu \circ \pi^{-1}$  with  $\pi : \mathbb{R} \rightarrow [-2\varepsilon, 2\varepsilon]$  being the projection onto the interval  $[-2\varepsilon, 2\varepsilon]$ . Then the exit times  $T$  and

$$T^* = \inf\{t \geq 0 : |X_t^*| \geq \varepsilon\}$$

are equal in law. Moreover, the process  $X_{T^* \wedge \cdot}^*$  is a by  $3\varepsilon$  uniformly bounded martingale and the quadratic variation process  $[X^*]$  of  $X^*$  is a subordinator with Doob-Meyer Decomposition

$$[X^*]_t = ([X^*]_t - 4\varepsilon^2 F_1(2\varepsilon)t) + 4\varepsilon^2 F_1(2\varepsilon)t.$$

Therefore,

$$\begin{aligned} 9\varepsilon^2 &\geq \mathbb{E}(X_{T^*}^2) = \lim_{t \rightarrow \infty} \mathbb{E}(X_{t \wedge T^*}^2) \\ &= \lim_{t \rightarrow \infty} \mathbb{E}[X]_{t \wedge T^*} = 4\varepsilon^2 F_1(2\varepsilon) \lim_{t \rightarrow \infty} \mathbb{E}(t \wedge T^*) = 4\varepsilon^2 F_1(2\varepsilon) \mathbb{E}(T^*). \end{aligned}$$

Consequently,

$$\mathbb{E}T^* \leq \frac{9}{4F_1(2\varepsilon)}$$

and the assertion follows immediately.  $\square$

**Lemma 3.8.** *Let  $Y$  be a Bernoulli r.v. Then for  $d \in [0, 1/2]$*

$$D(d \log 2d + (1-d) \log 2(1-d) \mid Y, \rho_{\text{Ham}}) \geq d,$$

where  $\rho_{\text{Ham}}$  denotes the Hamming distance.

**Proof.** Interpret  $Y$  as a random variable attaining values in the group  $\mathbb{Z}_2$  consisting of two elements. Then  $\rho$  can be interpreted as a difference distortion measure on  $\mathbb{Z}_2$ , that means for  $x, \hat{x} \in \mathbb{Z}_2$

$$\rho(x, \hat{x}) = \rho(x - \hat{x}) := \mathbb{1}_{\{x - \hat{x} = 0\}}.$$

Next, note that for  $d \in [0, 1/2]$ :

$$\phi(d) := \sup\{H(Z) : Z \text{ } \mathbb{Z}_2\text{-valued, } \mathbb{E}[\rho(Z)] \leq d\} = -d \log d - (1-d) \log(1-d).$$

We use the concept of the Shannon lower bound to finish the proof: Let  $\hat{Y}$  denote a  $\mathbb{Z}_2$ -valued reconstruction with  $\mathbb{E}[\rho(Y, \hat{Y})] = d \leq 1/2$ ; then

$$\begin{aligned} I(Y; \hat{Y}) &= H(Y) - H(\hat{Y} \mid Y) = H(Y) - H(\hat{Y} - Y \mid Y) \geq H(Y) - H(\hat{Y} - Y) \\ &\geq \log 2 - \phi(d) = d \log 2d + (1-d) \log 2(1-d). \end{aligned}$$

$\square$

In the proof we will use that for the Bernoulli distribution  $\mu^{\text{Ber}}$  and Hamming distortion  $\rho^{\text{Ham}}$  one has for any  $d \in [0, 1/2]$  that

$$D(d \log 2d + (1-d) \log 2(1-d) \mid \mu^{\text{Ber}}, \rho^{\text{Ham}}) = d.$$

The proof of the lower bound is based on a comparison with a simpler distortion rate function. For  $q \in [0, 1/2]$  let  $\mu_q$  denote the measure that assigns probabilities  $q$  to  $\pm 1$  and  $1 - 2q$  to 0. Moreover denote by  $\mu_q^{\otimes n}$  its product measure, consider the distortion measure

$$\rho(x, \hat{x}) = \mathbb{1}_{\{x \cdot \hat{x} = -1\}} \quad (x \in \{\pm 1, 0\}, \hat{x} \in \{\pm 1\})$$

and denote

$$\rho_n(x, \hat{x}) = \sum_{i=0}^{n-1} \rho(x_i, \hat{x}_i).$$

As reconstruction we allow any  $\{\pm 1\}^n$ -valued random vector.

**Proposition 3.9.** *For any  $r \geq 0$ ,  $n \in \mathbb{N}$  and any Lévy process with symmetric Lévy measure, one has*

$$D(r|\mathbb{P}_X, s) \geq \frac{\varepsilon}{4n^{1/p}} D(r|\mu_q^{\otimes n}, \rho_n, s).$$

where

$$q = \frac{1}{8} \left( 1 - \frac{9}{F_1(2\varepsilon)l} \right) \vee 0.$$

**Proof.** First fix  $n \in \mathbb{N}$ ,  $r \geq 0$  and a reconstruction  $\hat{X}$  with  $I(X; \hat{X}) \leq r$ . We denote  $l = 1/n$  and consider again

$$\pi : L_1[0, 1] \rightarrow \ell_1^n, \quad (x_t) \mapsto \left( \left| \int_{il}^{(i+1)l} (2\mathbf{1}_{\{t \geq (2i+1)l/2\}} - 1)x_t \frac{dt}{l} \right| \right)_{i=0, \dots, n-1}.$$

The map  $\pi$  is  $l^{-1}$ -Lipschitz continuous and the random vector

$$Y := (Y_i)_{i=0, \dots, n-1} = \pi(X)$$

consists of i.i.d. entries. Additionally, we set  $\hat{Y} = (\hat{Y}_i)_{i=0, \dots, n-1} = \pi(\hat{X})$ . Next, consider random vectors  $Z = (Z_i)_{i=0, \dots, n-1}$  and  $\hat{Z} = (\hat{Z}_i)_{i=0, \dots, n-1}$  defined as

$$Z_i = \begin{cases} \text{sgn}(Y_i) & \text{if } |Y_i| \geq \varepsilon/4 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{Z}_i = \begin{cases} 1 & \text{if } \hat{Y}_i \geq 0 \\ -1 & \text{otherwise.} \end{cases}$$

Recalling the Lipschitz continuity of  $\pi$  we get that

$$\|X - \hat{X}\| \geq l \|Y - \hat{Y}\|_{\ell_1^n} \geq l \frac{\varepsilon}{4} \sum_{i=0}^{n-1} \rho(Z_i, \hat{Z}_i).$$

Therefore,

$$\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} \geq \frac{\varepsilon}{4n} \mathbb{E}[\rho_n(Z, \hat{Z})^s]^{1/s}.$$

Certainly,  $Z$  is distributed according to  $\mu_q^{\otimes n}$ , where  $q = \mathbb{P}(Y_1 \geq \varepsilon/4)$ . Since  $I(X; \hat{X}) \geq I(Z; \hat{Z})$  we obtain that in general

$$D(r|\mathbb{P}_X, s) \geq \frac{\varepsilon}{4n^{1/p}} D(r|\mu_q^{\otimes n}, \rho_n, s).$$

Next, we show that  $D(r|\mu_q^{\otimes n}, \rho_n, s)$  is increasing in  $q$ . Indeed, let  $0 \leq q < q' \leq 1/2$ , let  $Z$  denote an  $\mu_{q'}^{\otimes n}$  distributed r.v., and let  $\hat{Z}$  denote a reconstruction for  $Z$  with  $I(Z; \hat{Z}) \leq r$ . Moreover, let  $A = (A_0, \dots, A_{n-1})$  be a random vector consisting of i.i.d. Bernoulli random variables with success probability  $q/q'$  that are independent of  $Z$  and  $\hat{Z}$  (for finding such a sequence one might need to enlarge the probability space), and set  $\tilde{Z} := (\tilde{Z}_i)_{i=0, \dots, n-1} := (A_i Z_i)_{i=0, \dots, n-1}$ . Then  $\tilde{Z}$  is  $\mu_q^{\otimes n}$ -distributed and one has

$$\mathbb{E}[\rho_n(\tilde{Z}, \hat{Z})] \leq \mathbb{E}[\rho_n(Z, \hat{Z})] \quad \text{and} \quad I(\tilde{Z}; \hat{Z}) \leq I(A, Z; \hat{Z}) = I(Z; \hat{Z}).$$

It remains to prove that  $\mathbb{P}(Y_i \geq \varepsilon/4) \geq \frac{1}{8} \left( 1 - \frac{9}{F_1(2\varepsilon)l} \right)$ . We fix  $i \in \{0, \dots, n-1\}$  and let

$$(\tilde{X}_t)_{t \in [-l/2, l/2]} = (X_{t + \frac{2i+1}{2}l} - X_{\frac{2i+1}{2}l})_{t \in [-l/2, l/2]}.$$

The processes  $(\tilde{X}_t)_{t \in [0, l/2]}$  and  $(-\tilde{X}_{-t})_{t \in [0, l/2]}$  are independent Lévy martingales with Lévy measure  $\nu$ . Denote  $T^+ = \inf\{t \geq 0 : \tilde{X}_t \geq \varepsilon \text{ or } t \geq l/2\}$  and observe that

$$\begin{aligned} \mathbb{P}(Y_i \geq \frac{\varepsilon}{4}) &\geq \mathbb{P}\left(-\int_0^{l/2} \tilde{X}_{-t} dt \geq 0, T \leq l/4, \int_T^{l/2} [\tilde{X}_t - \tilde{X}_T] dt \geq 0\right) \\ &= \mathbb{P}\left(\int_0^{l/2} \tilde{X}_{-t} dt \leq 0\right) \mathbb{P}(T \leq l/4) \mathbb{P}\left(\int_T^{l/2} [\tilde{X}_t - \tilde{X}_T] dt \geq 0 \mid T \leq l/4\right) \\ &= \frac{1}{4} \mathbb{P}(T^+ \leq l/4). \end{aligned}$$

Set  $T = \inf\{t \geq 0 : |\tilde{X}_t| \geq \varepsilon \text{ or } t \geq l/2\}$ . Then the symmetry of  $\nu$  together with Lemma 3.7 implies that

$$\mathbb{P}(T^+ \leq l/4) \geq \frac{1}{2} \mathbb{P}(T \leq l/4) \geq \frac{1}{2} \left(1 - \frac{9}{F_1(2\varepsilon)l}\right)$$

so that

$$\mathbb{P}(Y_i \geq \frac{\varepsilon}{4}) \geq \frac{1}{8} \left(1 - \frac{9}{F_1(2\varepsilon)l}\right).$$

□

**Lemma 3.10.** *Let  $\mu^{\text{Ber}}$  and  $\rho^{\text{Ham}}$  denote the Bernoulli distribution and the Hamming distance, respectively. Then*

$$D(r \mid \mu_q, \rho) \geq 2q D\left(\frac{r}{2q} \mid \mu^{\text{Ber}}, \rho^{\text{Ham}}\right).$$

**Proof.** Let  $X$  denote a  $\mu_q$  distributed r.v. and let  $\hat{X}$  denote a  $\{\pm 1\}$ -valued reconstruction with  $I(X; \hat{X}) \leq r$ . Denote  $f(\bar{x}) = I(X; \hat{X} \mid |X| = \bar{x})$  for  $\bar{x} \in \{0, 1\}$  and let

$$\bar{r} = f(1) \quad \text{and} \quad R = f(|X|).$$

Then one has  $\mathbb{E}R = I(X; \hat{X} \mid |X|) \leq I(X; \hat{X}) \leq r$  so that due to the non-negativity of  $R$

$$\bar{r} \leq \frac{r}{\mathbb{P}(|X| = 1)} = \frac{r}{2q}.$$

Next, we write

$$\mathbb{E}\rho(X, \hat{X}) = \mathbb{E}\left[\mathbb{1}_{\{X \neq 0\}} \mathbb{E}[\mathbb{1}_{\{X \neq \hat{X}\}} \mid |X|]\right]$$

and note that conditional on  $|X| = 1$ ,  $X$  is a Rademacher random variable so that

$$\mathbb{E}\rho(X, \hat{X}) \geq \mathbb{P}(|X| = 1) D(\bar{r} \mid \mu^{\text{Ber}}, \rho^{\text{Ham}}).$$

Together with the above estimate for  $\bar{r}$  this completes the proof. □

**Proof of Theorem 3.5, 1<sup>st</sup> statement.** Let  $\varepsilon > 0$  with  $F_1(2\varepsilon) \geq 18$  and choose  $n \in \mathbb{N}$  maximal with  $n \leq F_1(2\varepsilon)/18$ . Then

$$q := \frac{1}{8} \left(1 - \frac{9n}{F_1(2\varepsilon)}\right) \vee 0 \geq \frac{1}{16}.$$

Additionally, there exists a universal constant  $C_3 > 0$  such that  $n \geq C_3 F_1(2\varepsilon)$ . Next, we shall apply Proposition 3.9. We fix  $r_0 < \log 2$  arbitrarily and set  $r = \frac{1}{8}nr_0$ . Then  $r \geq C_1 F_1(2\varepsilon)$  for some constant  $C_1$  only depending on the choice of  $r_0$ . Thus with Proposition 3.9 one gets

$$D(C_1 F_1(2\varepsilon), s) \geq D(r, s) \geq \frac{\varepsilon}{4n} D\left(\frac{1}{8}nr_0 \middle| \mu_q^{\otimes n}, \rho_n, s\right). \quad (25)$$

Recall that statement 1 of the theorem considers the case where  $s = 1$ . But  $D\left(\frac{1}{8}nr_0 \middle| \mu_q^{\otimes n}, \rho_n\right)$  is a distortion rate function for a single letter distortion measure and an i.i.d. original, and, therefore,

$$D\left(\frac{1}{8}nr_0 \middle| \mu_q^{\otimes n}, \rho_n\right) = n D\left(\frac{1}{8}r_0 \middle| \mu_q, \rho\right)$$

The latter distortion rate function has been related to that of a Bernoulli variable in Lemma 3.10:

$$D\left(\frac{1}{8}nr_0 \middle| \mu_q^{\otimes n}, \rho_n\right) \geq n 2q D\left(\frac{1}{16q}r_0 \middle| \mu^{\text{Ber}}, \rho^{\text{Ham}}\right).$$

Since  $q \geq 1/16$  the rate in the last distortion rate function is bounded by  $r_0 < \log 2$  so that the distortion rate function yields a value  $C_4 > 0$  strictly bigger 0. Altogether,

$$D(C_1 F_1(2\varepsilon), 1) \geq \frac{\varepsilon}{2} q C_4 \geq C_2 2\varepsilon,$$

where  $C_2 = \frac{1}{8}(C_4/8)^{1/p}$ . Switching from  $2\varepsilon$  to  $\varepsilon$  finishes the proof of the first assertion.  $\square$

The proof of the second statement relies on the following concentration property:

**Lemma 3.11.** *Let  $\rho : \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty]$  be a measurable function, let  $(U_i)_{i \in \mathbb{N}}$  be a sequence of independent bounded random variables, and denote by  $U^{(n)}$  the random vector  $(U_i)_{i=1, \dots, n}$ . Supposing that there exists  $u^* \in \mathbb{R}$  such that*

$$\mathbb{E}[\rho(U_1, u^*)^2] < \infty, \quad (26)$$

one has for any  $s > 0$  and  $r > 0$ :

$$\liminf_{n \rightarrow \infty} \frac{1}{n} D(nr | U^{(n)}, \rho_n, s) \geq d,$$

where  $d = D(r | U_1, \rho, 1)$  and  $\rho_n$  is the single letter distortion measure belonging to  $\rho$ .

As one can see in the proof the moment condition (26) can be easily relaxed. Similar ideas are used in [3] to prove concentration of the approximation error.

**Proof.** Without loss of generality we assume that  $D(r | U_1, \rho) > 0$ . Our moment condition implies that  $D(\cdot | U_1, \rho)$  is finite, convex and continuous on  $[0, \infty)$ . Following the standard proof of Shannon's source coding theorem, there is a family of codebooks  $(\mathcal{C}(n))_{n \in \mathbb{N}}$  such that

- $\{(u^*, \dots, u^*)\} \subset \mathcal{C}(n) \subset \mathbb{R}^n$ ,
- $\log |\mathcal{C}(n)| \lesssim nr$ ,
- $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{T}(n)) = 1$  for  $\mathcal{T}(n) = \{\min_{\hat{u}^{(n)} \in \mathcal{C}(n)} \rho_n(U^{(n)}, \hat{u}^{(n)}) < (1 + \varepsilon(n))d\}$  and an appropriate zero-sequence  $(\varepsilon(n))_{n \in \mathbb{N}}$ .



For any  $n \in \mathbb{N}$ , let  $\hat{U}^{(n,1)}$  denote an arbitrary reconstruction for  $U^{(n)}$  such that we have  $I(U^{(n)}, \hat{U}^{(n,1)}) \leq nr$ , and let  $\hat{U}^{(n,2)} = \arg \min_{\hat{u}^{(n)} \in \mathcal{C}(n)} \rho_n(U^{(n)}, \hat{u}^{(n)})$ . We fix  $\eta \in (0, 1)$  arbitrarily and choose

$$J = \begin{cases} 1 & \text{if } \log \frac{d\mathbb{P}_{U^{(n)}, \hat{U}^{(n,1)}}}{d\mathbb{P}_{U^{(n)} \otimes \mathbb{P}_{\hat{U}^{(n,1)}}}} \leq nr \text{ and } \rho_n(U^{(n)}, \hat{U}^{(n,1)}) \leq (1 - \eta)d, \\ 2 & \text{else,} \end{cases}$$

and  $\hat{U}^{(n)} = \hat{U}^{(n,J)}$ .

Next, we will use that

$$I(U^{(n)}; \hat{U}^{(n)}) \leq I(U^{(n)}; \hat{U}^{(n)}, J) = \inf_Q H(\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J} \| \mathbb{P}_{U^{(n)}} \otimes Q),$$

where the infimum is taken over all probability measures  $Q$  on  $\mathbb{R} \times \{1, 2\}$  and  $H$  denotes the relative entropy. We choose

$$Q = \frac{1}{2} [\mathbb{P}_{\hat{U}^{(n,1)}} \otimes \delta_1 + Q^* \otimes \delta_2] \quad \text{with} \quad Q^* = \frac{1}{|\mathcal{C}(n)|} \sum_{\hat{u}^{(n)} \in \mathcal{C}(n)} \delta_{\hat{u}^{(n)}}$$

in order to get an appropriate bound for  $I(U^{(n)}; \hat{U}^{(n)})$ :

$$\begin{aligned} I(U^{(n)}; \hat{U}^{(n)}) &\leq H(\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J} \| \mathbb{P}_{U^{(n)}} \otimes Q) \\ &= \int \log \frac{d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J}}{d\mathbb{P}_{U^{(n)}} \otimes Q} d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J} \\ &\leq \int_{\{J=1\}} \log \frac{d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J}}{d\mathbb{P}_{U^{(n)}} \otimes \mathbb{P}_{\hat{U}^{(n,1)}} \otimes \delta_1} d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J} \\ &\quad + \int_{\{J=2\}} \log \frac{d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J}}{d\mathbb{P}_{U^{(n)}} \otimes Q^* \otimes \delta_2} d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J} + \log 2 \end{aligned}$$

Note that the measures  $\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J}$  and  $\mathbb{P}_{U^{(n)}, \hat{U}^{(n,1)}, J}$  agree on the set  $\{J = 1\}$  so that by the construction of  $J$  one has  $\log \frac{d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J}}{d\mathbb{P}_{U^{(n)}} \otimes \mathbb{P}_{\hat{U}^{(n,1)}} \otimes \delta_1} \leq nr$  on  $\{J = 1\}$ . Moreover, one has  $\log \frac{d\mathbb{P}_{U^{(n)}, \hat{U}^{(n)}, J}}{d\mathbb{P}_{U^{(n)}} \otimes Q^* \otimes \delta_2} \leq \log |\mathcal{C}(n)|$  on  $\{J = 2\}$ . Consequently, we can continue with

$$I(U^{(n)}; \hat{U}^{(n)}) \leq P(J = 1) nr + P(J = 2) \log |\mathcal{C}(n)| + \log 2 \lesssim nr.$$

On the other hand, basic transformations and the Cauchy-Schwarz Inequality yield

$$\begin{aligned} &\mathbb{E}[\rho_n(U^{(n)}, \hat{U}^{(n)})] \\ &= \mathbb{E}[\mathbb{1}_{\{J=1\}} \rho_n(U^{(n)}, \hat{U}^{(n,1)})] + \mathbb{E}[\mathbb{1}_{\{J=2\}} \rho_n(U^{(n)}, \hat{U}^{(n,2)})] \\ &\leq (1 - \eta) d \mathbb{P}(J = 1) + \mathbb{P}(J = 2) (1 + \varepsilon(n)) d + \mathbb{P}(\mathcal{T}^c)^{1/2} \mathbb{E}[\rho_n(U^{(n)}, (u^*, \dots, u^*))^2]^{1/2} \\ &\sim [(1 - \eta) \mathbb{P}(J = 1) + \mathbb{P}(J = 2)] d. \end{aligned}$$

Therefore,  $\lim_{n \rightarrow \infty} \mathbb{P}(J = 1) = 0$ . Consequently, we arrive at

$$\mathbb{E}[\rho_n(U^{(n)}, \hat{U}^{(n,1)})^s]^{1/s} \geq \mathbb{P}(J \neq 1)^{1/s} (1 - \eta) d \rightarrow (1 - \eta) d$$

and recalling that  $\eta \in (0, 1)$  was arbitrary finishes the proof.  $\square$

**Proof of Theorem 3.5, 2<sup>nd</sup> statement.** We define  $r_0$ ,  $q$  and  $n$  as in the proof of the first statement. By assumption  $\nu(\mathbb{R}) = \infty$  or  $\sigma \neq 0$ . Consequently, one has  $\lim_{\varepsilon \downarrow 0} F_1(\varepsilon) = \infty$  and  $n$  converges to  $\infty$  as  $\varepsilon$  tends to 0.

We recall estimate (25):

$$D(C_1 F_1(2\varepsilon), s) \geq D(r, s) \geq \frac{\varepsilon}{4n} D\left(\frac{1}{8}nr_0 \Big| \mu_q^{\otimes n}, \rho_n, s\right).$$

Now we conclude with Lemma 3.11 that

$$D\left(\frac{1}{8}nr_0 \Big| \mu_q^{\otimes n}, \rho_n, s\right) \gtrsim nD\left(\frac{1}{8}r_0 \Big| \mu_q, \rho\right).$$

The assertion follows along the lines of the proof of the first statement.  $\square$

## Acknowledgement

The research of the first-mentioned author was supported by the DFG Research Center “MATH-EON – Mathematics for key technologies” in Berlin.

## References

- [1] J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.
- [2] J. Creutzig. Approximation of S $\alpha$ S Lévy processes in  $L_p$  norm. *J. Approx. Theory*, 140(1):71–85, 2006.
- [3] S. Dereich. The coding complexity of diffusion processes under  $L^p[0, 1]$ -norm distortion. Preprint, 2006.
- [4] P. Elias. Universal codeword sets and representations of the integers. *IEEE Trans. Inform. Theory*, 21:194–203, 1975.
- [5] S. Ihara. *Information theory for continuous systems*. Singapore: World Scientific, 1993.
- [6] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Internat. J. Comput. Math.*, 2:157–168, 1968.
- [7] H. Luschgy and G. Pagès. Functional quantization rate and mean pathwise regularity of processes with an application to Lévy processes. 2007. Preprint.
- [8] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2004.
- [9] C. Vormoor. *High resolution coding of point processes and the Boolean model*. PhD thesis, Technische Universität Berlin, 2007.