The coding complexity of Lévy processes

by

Frank Aurzada and Steffen Dereich

Communicated by Henryk Wozniakowski

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

July 17, 2008

Summary. We introduce a new coding scheme for general real-valued Lévy processes and control its performance w.r.t. $L_p[0,1]$-norm distortion under different complexity contraints. We also establish lower bounds that prove the optimality of our coding scheme in many cases.

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

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

1 Introduction and results

1.1 Motivation

The objective of the article is

(I) to provide efficient coding strategies together with error bounds for general Lévy processes and

(II) to complement the error bounds by appropriate lower bounds that show weak optimality of our scheme for most cases.

Let us be more precise. We study the coding problem for real-valued Lévy processes X (the *original* or source signal) 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 $0 < s \le \infty$. The objective is to find a càdlàg, real-valued process $\hat{X}$ (*reconstruction* or *approximation*) that minimizes the error criterion

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

under a given complexity constraint on the approximating random variable $\hat{X}$. We work with the following three classical complexity constraints (see for instance Kolmogorov [14]): for $r \geqslant 0$,

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

We 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

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

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

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

$$D^{(q)}(r,s) := \inf\big\{\big\|\|X-\hat{X}\|\big\|_{L_s(\mathbb{P})} : \log|\operatorname{range}(\hat{X})| \le r\big\},$$

which we call the (minimal) *quantization error* for the *rate* $r \ge 0$ 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. The quantities $D^{(e)}$ and D are called entropy coding error and distortion rate function, respectively. In general, $I(X;\hat{X}) \leqslant H(\hat{X}) \leqslant \log|\operatorname{range}(\hat{X})|$ so that we have $D \leqslant D^{(e)} \leqslant D^{(q)}$, cf. e.g. [4], Section 2.4. In the following, D, $D^{(e)}$, and $D^{(q)}$ are called *coding quantities*. Strictly speaking, the quantities $D^{(e)}$ and D depend on the probability space. However, our results are valid independently of the choice of the probability space.

The quantization constraint naturally appears, when coding the signal X under a strict bit-length constraint for its binary representation. On the other hand, the entropy constraint corresponds to an average bit-length constraint and optimal codes are obtained via Huffman coding. Allowing simultaneous coding of several independent copies of the source signal and measuring the error by the sum of the individual errors results in a further increase in the efficieny. Due to Shannon's source coding theorem, the distortion-rate function D measures the corresponding best-achievable error. A more recent motivation for studying quantization is its applicability in numerical integration, see e.g. [18], [9].

In general, one is interested in the asymptotic behavior of the above quantities (which may serve as benchmarks) and in algorithms that are close to optimal. If the source signal lies in an infinite dimensional Banach space, we speak of the *functional* coding problem. Functional coding, and in particular functional quantization, is intensively studied since the beginning of this century. For Gaussian measures in Hilbert spaces the problem is well understood, see [15], [6]. In Banach spaces, the problem is closely related to the theory of small deviations and it is possible to deduce the weak asymptotics in many cases, see [10]. For fractional Brownian motion and for one-dimensional diffusion processes, the asymptotic behavior of the coding quantities is established in [11], [7], and [8]; see also [16] for further, constructive results. In general, good approximation schemes can often be implemented by using series expansions for the signal together with appropriately adjusted quantizers for the coefficients, see [5], [17].

Next, we shall introduce the main notation. Then in Section 1.3 it follows the statement of the main asymptotic results. Section 1.4 introduces the central coding scheme used to achieve task (I), and Section 1.5 proceeds with a list of applications to important examples of Lévy processes.

Section 2 is devoted to the proof of the upper bounds needed for task (I), and finally the proof of the corresponding lower bounds – showing optimality in most cases, i.e. (II) – is provided in Section 3.

1.2 Main notation

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, cf. [2, 20]. 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

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

for parameters $\sigma^2 \in [0,\infty)$, $b \in \mathbb{R}$, and a positive measure ν on $\mathbb{R}$ that has no point mass at zero and satisfies

$$\int_{\mathbb{R}} 1 \wedge x^2\,\nu(dx) < \infty, \tag{3}$$

where we use $x\wedge y := \min(x,y)$ for $x, y \geqslant 0$. On the other hand, for a given triplet (ν, σ^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 call the corresponding process an (ν, σ^2, b)-Lévy process.

If (2) is true for

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

then we call X an (ν, σ^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.

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

1.3 Results

The crucial terms determining the behavior of the coding quantities of Lévy processes are

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

Furthermore, we set $F(\varepsilon) = F_1(\varepsilon)+F_2(\varepsilon)$. Note that $F(\varepsilon)$ is obtained by integrating the function visualised in Figure 1 with respect to the Lévy measure. Furthermore, observe that (3) does not ensure the finiteness of F_2 and that F_2 is either finite or infinite for all $\varepsilon > 0$.

Let us start with our main result for the entropy coding error.

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 $\varepsilon \in (0, 1]$,*

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

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) $\mathbb{E}\,\|X\|^q < \infty$ *and*

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

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

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

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

In the proofs of the upper bounds we only need to consider the case where F_2 is finite. Indeed, assumption (a) in Theorem 1.2 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. 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 ε tends to zero. We note here that some condition of this kind is necessary (though probably (b) can be softened), since for finite Lévy measure the rate is different to the one that could be expected from Theorem 1.2, cf. remarks in Example 1.11 and [1]. With given Lévy measure, it is usually easy to verify conditions (a) and (b), cf. Remarks 2.2 and 2.3 below.

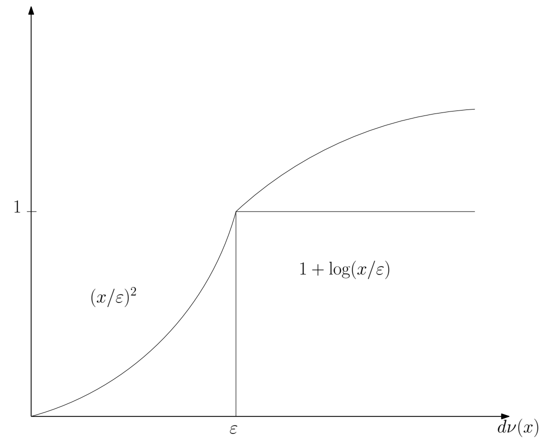

Figure 1: Visualization of the function F

Remark 1.4. Alternatively, the complexity of a signal can also be measured in terms of the best finite-dimensional approximation (Kolmogorov widths). This approach was investigated for symmetric α-stable Lévy processes in [5] and used for generating quantization schemes. These schemes are optimal if and only if $p < \alpha$.

A similar approach for the quantization of Lévy processes is taken in [17]. There, linear quantizers are constructed from a series representation. Then the path regularity of the process allows to derive error bounds for the approximation. The results generalize those from [5] to further stable-like Lévy processes and for p smaller than the so-called Blumenthal-Getoor index.

In this article, we work with non-linear quantizers, which lead to weakly optimal results in most cases. We compare the results to those from [5] and [17] when looking at the examples in Section 1.5. Our lower bounds also show the optimality of the results from [5] and [17] for small p.

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 $F_1(\varepsilon) \geqslant c_3$ one has*

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

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 ε 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.

Remark 1.7. Heuristically, one can explain the appearance of F_1 and F_2 in the theorems as follows. For jumps that are significantly larger than ε one needs to provide sufficiently accurate information about location and height of the jump in order to recover the process up to an error of order ε. This induces the cost term $F_1(\varepsilon)$. On the other hand, the small jumps are compared with their mean behavior, and significant deviations appear in intervals of length $1/F_2(\varepsilon)$. This results in the cost term $F_2(\varepsilon)$.

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 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.

1.4 An explicit coding strategy

We now introduce the central coding scheme. In the first step, we define the approximation $\hat{X}$ to X, and in the second step we describe how to get an appropriate binary representation for $\hat{X}$. The coding scheme depends on a parameter $\varepsilon \in (0,1]$ that is fixed in the following discussion and that indicates the accuracy we are aiming at.

Approximating terms

First denote by $(X'_t)_{t\geqslant 0}$ the process given by $X'_t = X_t - b_\varepsilon t$, where

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

We use projections on the $\varepsilon\mathbb{Z}$ grid to approximate X'. For this purpose, let us define g to be the (right-continuous) nearest neighbor projection of $\mathbb{R}$ onto $\varepsilon\mathbb{Z}$, and define inductively stopping times $(S_i)_{i\in\mathbb{N}_0}$ as follows: set $S_0 = 0$ and, for $i \in \mathbb{N}$, let us define the exit times S_i by

$$S_i := \inf\left\{t > S_{i-1} : |X'_t - g(X'_{S_{i-1}+})| \geqslant 2\varepsilon\right\}.$$

Moreover, let $M := \max\{i : S_i < 1\}$.

As an intermediate approximation to X', we use $\tilde{X}'$ defined by $\tilde{X}'_t = g(X'_{S_i+})$ for $t \in [S_i, S_{i+1})$ and $i \in \mathbb{N}_0$. Here we infer a loss: $\|X' - \tilde{X}'\| \leqslant 2\varepsilon$. Our coding scheme is based on encoding the jump times (S_i) and the jump heights $(H_i) := (\Delta\tilde{X}'_{S_i})$, where we denote as usual for any time $t > 0$ the jump of a process X at time t by $\Delta X_t = X_t - X_{t-}$. Whereas all jump heights occuring in $[0,1)$ can be encoded exactly, we need to work with approximations $(\hat{S}_i)$ for the relevant jump times (S_i). Here we demand that

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

Based on $(\hat{S}_i)$ and (H_i) we approximate X' and X by

$$\hat{X}'_t = \sum_{i=1}^{M} H_i \mathbb{1}_{[\hat{S}_i,\infty)}(t) \qquad \text{and} \qquad \hat{X}_t = \hat{X}'_t + b_\varepsilon t,$$

respectively. Once we have established appropriate coding schemes for (S_i) and (H_i) ensuring (5), we thus can recover the original X by $\hat{X}$ with an error smaller or equal to 3ε. Indeed,

$$\|X - \hat{X}\| = \|X' - \hat{X}'\| \leqslant \|X' - \tilde{X}'\| + \|\tilde{X}' - \hat{X}'\| \leqslant 2\varepsilon + \Big[\sum_{i=1}^{M} |H_i|^p(\hat{S}_i - S_i)\Big]^{1/p} \leqslant 3\varepsilon. \tag{6}$$

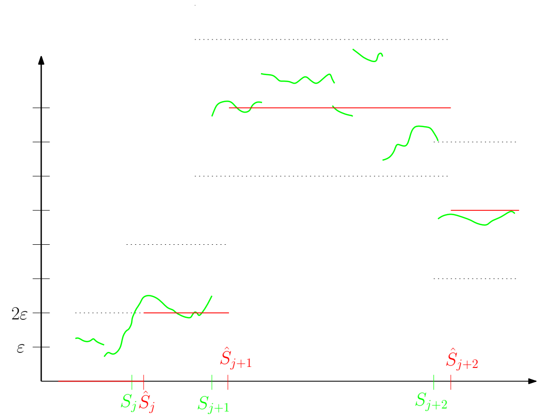

Figure 2: The coding procedure

Coding procedure

We use binary prefix-free representations Υ_1 and Υ_2 for $\mathbb{Z}$ and $\mathcal{D} := [0,1) \cap \bigcup_{n\in\mathbb{N}} 2^{-n}\mathbb{Z}$, respectively. For our purposes, it is sufficient that these prefix-free representations satisfy for all $m \in \mathbb{Z}$

$$\text{length}\, \Upsilon_1(m) \leqslant \text{const}\ (1 + \log_+ |m|)$$

and for all $n \in \mathbb{N}$ and $x \in [0,1] \cap 2^{-n}\mathbb{Z}$

$$\text{length}\, \Upsilon_2(x) \leqslant \text{const}\ n.$$

Though this might be well known, we provide an explicit construction in Lemmas 2.5 and 2.6, respectively.

Each jump height H_i is translated into the code

$$\pi_i^H := \Upsilon_1(H_i/\varepsilon).$$

The coding of (S_i) is more intricate: we divide the interval $[0,1)$ into boxes (i.e. intervals) $I_j = [jF_1(\varepsilon)^{-1}, (j+1)F_1(\varepsilon)^{-1} \wedge 1)$ $(j = 0, \dots, \lceil F_1(\varepsilon)\rceil - 1)$. Then a jump time S_i in box I_j corresponds to a value

$$\frac{S_i - jF_1(\varepsilon)^{-1}}{F_1(\varepsilon)^{-1}} \in [0,1).$$

Note that $\mathcal{D}$ contains only numbers with finite binary representation. We approximate S_i by $\hat{S}_i$ such that (5) is valid and such that

$$\frac{\hat{S}_i - jF_1(\varepsilon)^{-1}}{F_1(\varepsilon)^{-1}} \in \mathcal{D} \tag{7}$$

has shortest binary representation. This leads to the code

$$\pi_i^S := \Upsilon_2\Big(\frac{\hat{S}_i - jF_1(\varepsilon)^{-1}}{F_1(\varepsilon)^{-1}}\Big).$$

Note that given the box of S_i and π_i^S, one can recover $\hat{S}_i$ perfectly. Now we encode the pair $(H_i, \hat{S}_i)$ by

$$\pi_i := \text{`0'} * \pi_i^H * \pi_i^S,$$

and we encode each block I_j by

$$\Pi_j := \prod_{\substack{i\in\mathbb{N}\\ S_i\in I_j}} \pi_i.$$

Here, $*$ and $\prod$ denote the concatenation of binary strings of finite lengths. If the index set is empty in the definition of Π_j, then Π_j is asssumed to be the empty word. Finally, we describe the approximation $\hat{X}$ by

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

Let us now show how to recover the relevant (H_i) and $(\hat{S}_i)$ from a binary string $\tilde{\Pi}$ with prefix Π. First we set $\tilde{\Pi} = \Pi$ and $j = 0$. Then, as long as $j < \lceil 1/F_1(\varepsilon)\rceil$, we remove the first digit of $\tilde{\Pi}$ and carry out one of the following operations depending on its value:

- ‘0’: recover the *height* value of a jump by applying a decoder of Υ_1, remove the corresponding digits from $\tilde{\Pi}$, and then recover the corresponding *time* approximation by applying a decoder of Υ_2 and by considering that the time lies in the box I_j, again remove the corresponding digits from $\tilde{\Pi}$;
- ‘1’: increase j by 1.

1.5 Examples

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

Example 1.8 (Stable Lévy process)**.** Let us consider the case of an α-stable Lévy process with $0 < \alpha < 2$. 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 $F_1(\varepsilon) = \tilde{C}_1\varepsilon^{-\alpha}$ and $F_2(\varepsilon) = \tilde{C}_2\varepsilon^{-\alpha}$. 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 $p \geqslant 1$,

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

This improves the findings from [5] and [17], where the result for the quantization error was obtained for $p < \alpha$. We remark that the lower bound actually already follows from [10].

Note that the coding complexity of an α-stable Lévy process $(0 < \alpha < 2)$ 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 1.9 (Lévy process with non-vanishing Gaussian component)**.** It is easy to calculate that $F_i(\varepsilon) \leqslant c\varepsilon^{-2}$ 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,

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

and

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

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

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

and,

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

In fact, by a modification of (12) one can show that (b) is not necessary if $\sigma \neq 0$.

Example 1.10 (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 $\varepsilon \geqslant 0$

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

and

$$D(c_1'(\log 1/\varepsilon)^2, s) \geqslant c_2' \varepsilon.$$

Therefore,

$$D(r,s) = \exp\bigl(-e^{\mathcal{O}(1)}\sqrt{r}\bigr) \qquad \text{and} \qquad D^{(e)}(r,s) = \exp\bigl(-e^{\mathcal{O}(1)}\sqrt{r}\bigr). \tag{8}$$

Note that Theorem 1.2 does not apply since condition (4) fails to hold. For the quantization error much less is known: Note that the lower bound provided by $D^{(e)} \leqslant D^{(q)}$ in (8) and the upper bounds in [17], where it is shown that

$$D^{(q)}(r,s) \leqslant r^{-1/p+o(1)}, \qquad \text{for all } p > 0 \text{ and } 0 < s \leqslant p,$$

strongly differ.

Example 1.11 (Compound Poisson process)**.** Let $(N(t))_{t\geqslant 0}$ 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 $b = \mathbb{E}[Y\,\mathbb{1}_{\{|Y|\leqslant 1\}}]$.

It is immediately clear that $F_1(\varepsilon) \leqslant 1$ and $F_2(\varepsilon) \approx \mathbb{E}\left[\log\left(\frac{|Y|}{\varepsilon}\right) \mathbb{1}_{\{|Y|\geqslant \varepsilon\}}\right]$. Except for the trivial case $\nu = 0$ this behaves as $\log(1/\varepsilon)$ so that F_2 dominates F, when ε 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

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

and

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

Hence,

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

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.

The quantization error for these cases was investigated in [17], [21], and [1]. In particular, it can be shown that the rates of quantization and entropy coding error *differ*, which is a behavior that we cannot analyze with the methods in this article.

2 Upper bounds

2.1 Preliminary considerations concerning the coding scheme

We already obeserved in (6) that for given $\varepsilon \in (0,1]$, the approximation $\hat{X}$ defined in Section 1.4 satisfies $\|X - \hat{X}\| \leqslant 3\varepsilon$.

In this section we provide upper bounds for the number of bits needed to encode $\hat{X}$ by our coding scheme.

Proposition 2.1. *There exist independent uniformly distributed random variables (U_i) (on a sufficiently large probability space) such that for some constant $K = K(p)$ depending on p only,*

$$\operatorname{length}\Pi \;\leqslant\; K\sum_{i=1}^{N}\Bigl(1+\log\frac{1}{U_i}\Bigr) + K\sum_{t\in[0,1)}\log_+\frac{|\Delta X_t|}{\varepsilon} + F_1(\varepsilon)+1, \tag{9}$$

where $N := \min\{n \in \mathbb{N}_0 : \sum_{i=1}^{n} U_i > F_1(\varepsilon)\}$.

We remark that the (U_i) occuring in the proposition are coupled with the exit times (S_i).

Proof. In a first step we analyze the waiting time between two consecutive jump times of (S_i). For this purpose, let $X^{(1)}$ be the process consisting of the (finitely many) jumps of X' that are greater than ε and set $X^{(2)} := X' - X^{(1)}$. Note that $X^{(2)}$ is a $(\nu|_{[-\varepsilon,\varepsilon]}, \sigma^2)$-Lévy martingale. Denote by Γ_1 the stopping time induced by the first jump of $X^{(1)}$. Note that $|X'_{S_i} - g(X'_{S_i+})| \leqslant \varepsilon/2$ so that due to the strong Markov property one has for all $t \geqslant 0$,

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

where the last step is justified by Doob's martingale inequality. By the compensation formula ([2], p. 7) the last term equals $F_1(\varepsilon)t$.

Let $(U_i)_{i\geqslant 1}$ be independent random variables that are uniformly distributed on $[0,1]$. Then we have shown that for all jumps

$$\mathbb{P}\left(S_{i+1}-S_i \leqslant t \mid \mathcal{F}_{S_i}\right) \;\leqslant\; \min(tF_1(\varepsilon),1) = \mathbb{P}\left(U_i \leqslant F_1(\varepsilon)t\right),$$

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

$$F_1(\varepsilon)(S_{i+1} - S_i) \geqslant U_i. \tag{10}$$

In particular, we have $N \geqslant M$, for N as defined in the proposition.

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 Υ_1 needs less than

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

  bits by Lemma 2.5.
- 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 by Lemma 2.7.

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

$$2\sum_{i=1}^{M}\left[\log \frac{|H_i|}{\varepsilon} + \log_+ \frac{F_1(\varepsilon)^{-1}}{\varepsilon^p/(M|H_i|^p) \wedge (S_{i+1} - S_i)}\right] + 9M + \lceil F_1(\varepsilon) \rceil.$$

This equals (using $x \vee y := \max(x, y)$ for $x, y \geqslant 0$)

$$2\sum_{i=1}^{M}\left[\log \frac{|H_i|}{\varepsilon} + \log_+ \left(\frac{M|H_i|^p}{F_1(\varepsilon)\varepsilon^p} \vee \frac{1}{F_1(\varepsilon)(S_{i+1} - S_i)}\right)\right] + 9M + \lceil F_1(\varepsilon) \rceil.$$

By (10) and the inequality $\log_+(x \vee y) \leqslant \log_+ x + \log_+ y$, the latter is less than

$$2\sum_{i=1}^{M}\left[(1+p)\log_+ \frac{|H_i|}{\varepsilon} + \log_+ \frac{1}{U_i}\right] + 2M \log_+ \frac{M}{F_1(\varepsilon)} + 9M + \lceil F_1(\varepsilon) \rceil. \tag{11}$$

Next, recall from (10) that $F_1(\varepsilon)(S_i - S_{i-1}) \geqslant U_i$ so that

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

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

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

Thus we conclude with (11) that

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

Note that we can estimate $|H_i| \leqslant |\Delta X_{S_i}| + \frac{5}{2}\varepsilon$, so that basic analysis gives

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

Consequently, the bit-length is bounded as follows:

$$\operatorname{length}\Pi \leqslant 4\sum_{i=1}^{M}\log\frac{1}{U_i}+2(1+p)\sum_{t\in[0,1)}\log_{+}\frac{|\Delta X_t|}{\varepsilon}+(19+10p)M+\lceil F_1(\varepsilon)\rceil,$$

which implies the assertion. □

2.2 Proof of Theorem 1.1

Proof. By (6) the error is less than 3ε a.s. (and thus is the mean error, for all moments $s > 0$).

We use the coupling introduced in Proposition 2.1 and estimate

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

by the compensation formula ([19], p. 29). Furthermore, by Lemma 2.4, we have for some universal constant c,

$$\mathbb{E}\left[\sum_{i=1}^{N}(1+\log U_i^{-1})\right] \leqslant c(F_1(\varepsilon)+1).$$

Therefore, as asserted,

$$H(\hat{X}) \leqslant \mathbb{E}\big[\operatorname{length}\Pi\big] \leqslant K\,\mathbb{E}\Big[\sum_{i=1}^{N}\Big(1+\log\frac{1}{U_i}\Big)+\sum_{t\in[0,1)}\log_{+}\frac{|\Delta X_t|}{\varepsilon}\Big]+F_1(\varepsilon)+1 \leqslant c_1F(\varepsilon)+c_1.$$

□

2.3 Proof of Theorem 1.2

Proof. We use the coding scheme introduced in Section 1.4. We shall use Π as representation, whenever $\operatorname{length}\Pi \leqslant c_1\,F(\varepsilon)$, where c_1 is an appropriate constant chosen below. This is what we define to be the event $\mathcal{T}$, which we also call the 'typical case'. Otherwise, we encode X by the zero function. Then trivially $\log|\operatorname{range}(\hat{X})| \leqslant c_1\,F(\varepsilon)$.

Note that, by the exponential compensation formula ([2], p. 8),

$$\mathbb{P}\left(\sum_{t\in[0,1)} \log_+ \frac{|\Delta X_t|}{\varepsilon} > C_2 F(\varepsilon)\right) \leqslant \exp\left(-C_2\mu F(\varepsilon)\right)\mathbb{E}\left[\exp\left(\mu \sum_{t\in[0,1)} \log_+ \frac{|\Delta X_t|}{\varepsilon}\right)\right]$$
$$\leqslant \exp\left(-C_2\mu F(\varepsilon) - \int_{|x|\,\geqslant\,\varepsilon} 1-(|x|/\varepsilon)^{\mu}\,\nu(dx)\right) \leqslant \exp\left(-C_2\mu F(\varepsilon)+C_3F(\varepsilon)\right) \leqslant e^{-\frac{C_2}{2}\mu F(\varepsilon)}, \tag{12}$$

where C_3 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,

$$\mathbb{P}\left(\sum_{i=1}^{\lfloor C_1F(\varepsilon)\rfloor} (1+\log U_i^{-1}) > C_2F(\varepsilon)\right) \leqslant e^{-\frac{C_2}{2}\,F(\varepsilon)},$$

for C_2 large enough. Finally, one proves, e.g. using the same discretization as in (14), that for C_1 large enough,

$$\mathbb{P}\left(N > C_1F(\varepsilon)\right) \leqslant e^{-\frac{C_1}{2}\,F(\varepsilon)}.$$

For c_1 chosen appropriately (depending on C_1 and C_2), we conclude from (9) that for some positive constant C (depending on C_1 and C_2), we have

$$\mathbb{P}\left(\mathcal{T}^c\right) = \mathbb{P}\left(\operatorname{length}\Pi > c_1F(\varepsilon)\right) \leqslant e^{-\frac{C_2}{2}\,F(\varepsilon)} + e^{-\frac{C_1}{2}\,F(\varepsilon)} + e^{-\frac{C_1}{2}\,F(\varepsilon)} \leqslant e^{-CF(\varepsilon)}.$$

Let $r>0$ be chosen such that $1/q+1/r=1/s$. Let $0<\kappa<1$ be chosen small enough such that $C\nu([-\kappa,\kappa]^c)/2 \geqslant r$. This is possible, since $\nu([-\kappa,\kappa]^c)$ tends to infinity when $\kappa\to 0$, by condition (b). Then, for $\varepsilon < \kappa^2$,

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

Thus,

$$\mathbb{P}\left(\mathcal{T}^c\right) \leqslant e^{-C\,F(\varepsilon)} \leqslant \varepsilon^r. \tag{13}$$

We have for the mean error, using the Hölder Inequality and $s\geqslant 1$,

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

where the term in brackets is bounded, by assumption (a) and (13). Note that the argument works analogously for $0<s<1$. □

Remark 2.2. It is easy to see that condition (a) is equivalent to the condition

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

Remark 2.3. A sufficient condition for (b) to hold is that (a) holds and that $\nu([-2\varepsilon, 2\varepsilon]^c) \leqslant c \cdot \nu([-\varepsilon, \varepsilon]^c)$ for some $0 < c < 1$ and all $0 < \varepsilon \leqslant \varepsilon_0$. This can be seen as follows. Firstly, note that the assumption implies that $\varepsilon \mapsto \nu([-\varepsilon, \varepsilon]^c)$ increases at least polynomially when $\varepsilon \to 0$, i.e. $\nu([-\varepsilon, \varepsilon]^c) \geqslant K(\varepsilon_0, c)\varepsilon^{-h}$ with $h = (-\log c)/\log 2$ and some constant $K(\varepsilon_0, c) > 0$ depending on ε_0 and c. Secondly,

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

Choosing $0 < \mu < ((-\log c)/\log 2) \wedge q$ yields

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

with a constant $K(\mu, c) > 0$ only depending on μ and c. This implies (b) since $\mu < h = (-\log c)/\log 2$.

Note that, in particular, the assumption is satisfied 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 2.4. *Let $\lambda > 0$ and let $(U_i)_{i\geqslant 1}$ be an i.i.d. sequence of random variables uniformly distributed in $[0, 1]$. For $N := \min\{n \in \mathbb{N}_0 : \sum_{i=1}^n U_i > \lambda\}$ one has*

$$\mathbb{E}\left[\sum_{i=1}^{N}(1 + \log U_i^{-1})\right] \leqslant 12\lceil\lambda\rceil.$$

Proof. Let $s \in \mathbb{R}$, define $N(s) := \min\{n \in \mathbb{N}_0 : \sum_{i=1}^n U_i > s\}$, and consider the function

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

We are interested in $\Psi(\lambda)$. Clearly, $\Psi(s) = 0$ for $s \leqslant 0$ and Ψ is increasing. Moreover, one has for $s > 0$ and a random variable U uniformly distributed in $[0, 1]$,

$$\begin{aligned}\Psi(s) &= \int_0^{s\wedge 1}\left(1 + \log x^{-1} + \mathbb{E}\left[\sum_{i=1}^{N(s-x)}(1 + \log U_i^{-1})\right) d\mathbb{P}_U(x)\right] \\ &\leqslant 1 + \int_0^1 -\log x\,dx + \int_0^{s\wedge 1} \mathbb{E}\left[\sum_{i=1}^{N(s-x)}(1 + \log U_i^{-1})\,d\mathbb{P}_U(x)\right] \\ &= 1 + \log e + \int_0^{s\wedge 1}\Psi(s-x)\,d\mathbb{P}_U(x) \leqslant 3 + \int_0^{s\wedge 1}\Psi(s-x)\,d\mathbb{P}_U(x).\end{aligned}$$

Let us define

$$U' := \begin{cases} 0 & U \leqslant 1/2 \\ 1/2 & U > 1/2. \end{cases} \tag{14}$$

Then $U' \leqslant U$; and since Ψ is increasing, we have

$$\Psi(s) \leqslant 3 + \int_0^{s\wedge 1} \Psi(s-x)\,\mathbb{P}_{U'}(x) = 3 + \frac{1}{2}\,\Psi(s) + \frac{1}{2}\,\Psi\left(s - \frac{1}{2}\right).$$

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

$$\Psi(\lambda) \leqslant 6 + \Psi\left(\lambda - \frac{1}{2}\right) \leqslant 6 + 6 + \Psi(\lambda - 1) \leqslant \ldots \leqslant 12 \cdot \lceil \lambda \rceil.$$

□

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

Lemma 2.5. *There is a universal coding scheme that returns a prefix-free code $\Upsilon_1(x) \in \{0,1\}^*$ for a given integer $x \in \mathbb{Z}$, such that* $\operatorname{length}\Upsilon_1(x) \leqslant 2(2 + \log_+ x)$.

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 $2^{n-1} \leqslant x < 2^n$. 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 $2(1 + \log_+ x)$. □

Let us remark that Lemma 2.5 can be improved up to the order $\log_+ x + C \log_+ \log_+ x + D$, as shown in [12].

Let as above $\mathcal{D} := [0,1) \cap \bigcup_{n\in\mathbb{N}} 2^{-n}\mathbb{Z}$ be the set of those numbers in $[0,1)$ that have finite binary representation.

Lemma 2.6. *There exists a universal coding strategy $\Upsilon_2 : \mathcal{D} \to \{0,1\}^*$ that returns a prefix-free code such that, for any $n \in \mathbb{N}_0$ and $x \in 2^{-n}\mathbb{Z}$, we have* $\operatorname{length}\Upsilon_2(x) \leqslant 2(2+n)$.

Proof. Any number $x \in \mathcal{D}$ has a unique finite representation $x = k2^{-n}$, with k uneven, $1 \leqslant k \leqslant 2^n - 1$, $n \in \mathbb{N}$. As a prefix-free code $\Upsilon_2(x)$ we chose the prefix-free code for the integer $2^{n-1} + (k+1)/2$. If $x \in [0,1) \cap 2^{-n}\mathbb{Z}$, by Lemma 2.5, the length of $\Upsilon_1(2^{n-1} + (k+1)/2)$ is bounded by $2(1+n)$. □

The last lemma, in particular, implies the following for our coding scheme.

Lemma 2.7. *There exists a universal coding strategy that encodes the numbers S_i by $\hat{S}_i$ satisfying (5) and (7) such that the binary representation used to encode $\hat{S}_i$ has a length of at most*

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

Proof. Let n be the smallest integer such that

$$2^{-(n+1)} \leqslant \min\left(\frac{S_{i+1}-S_i}{F_1(\varepsilon)^{-1}}, \frac{\varepsilon^p/(|H_i|^p M)}{F_1(\varepsilon)^{-1}}\right) < 2^{-n}.$$

Then $\hat{S}_i$ can be chosen such that the number in (7) is in $2^{-n}\mathbb{Z}$ and that (5) is satisfied. According to Lemma 2.6 the length of the binary representation is bounded by $2(1+n)$, which is bounded by the asserted quantity, 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 prove some preliminary results. Next, we provide a lower bound based on F_2 (Theorem 3.3). In the last subsection we give a lower bound in terms of F_1 (Theorem 3.5). Both lower bounds then immediately imply Theorem 1.5.

So far, p was fixed 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 introduce some concepts of information theory. We need the concept of conditional mutual information. Let A, B and C denote random vectors attaining values in some standard 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

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

A summary of computation rules for the mutual information can be found in [13] or [4].

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 standard Borel spaces. We write shortly $Y=(Y_0,\dots,Y_{n-1})$, $Y^i=(Y_0,\dots,Y_i)$ for $0\leqslant i\leqslant n-1$ and $\hat{Y}=(\hat{Y}_0,\dots,\hat{Y}_{n-1})$. Then one has*

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

Moreover, we 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 μ on a standard Borel space E and a measurable function $\rho : E\times E\to[0,\infty]$ (*distortion measure*) we write

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

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

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

Moreover, we omit ρ 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 on the coding complexity of the Lévy process:

Proposition 3.2. *Let E be a standard 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 $\varphi : E \to E^2$ with*

$$\varphi(A+B) = (A,B) \ \ a.s. \tag{15}$$

Then, under any difference distortion measure ρ on E, one has for every $r \geqslant 0$:

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

Proof. Fix $r \geqslant 0$. 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, using the above mentioned convexity

$$\begin{aligned}\mathbb{E}[\rho(A+B,Z)] &= \int \mathbb{E}[\rho(A,Z-b)|B=b]\,d\mathbb{P}_B(b)\\ &\geqslant \int f(I(A;Z|B=b))\,d\mathbb{P}_B(b) = f\Big(\int I(A;Z|B=b)\,d\mathbb{P}_B(b)\Big)\\ &= f(I(A;Z|B)).\end{aligned}$$

Therefore,

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

On the other hand, by assumption (15), $I(A+B;Z) = I((A,B);Z)$ for any random element Z on E. Hence,

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

Therefore,

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

$\square$

3.2 Lower bound based on F_2

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

$$D\Big(\frac{\kappa(\varepsilon)}{e}\,F_2(\varepsilon)\Big|\mathbb{P}_X, L_1[0,1], 1\Big) \geqslant c\,\kappa(\varepsilon)\,\varepsilon,$$

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 ε, one needs to allocate about $\log_+ |X_t - X_{t-}|/\varepsilon$ bits for each big jump.

The problem is related to a minimization problem that we want to introduce now. Let Π be a finite non-negative measure on a measurable space $(E,\mathcal{E})$ and let $h : E \to [0,\infty)$ 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 $\xi : E \to [0,\infty)$ satisfying the constraint

$$\int \xi(x)\, d\Pi(x) \leqslant r. \tag{16}$$

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

$$\xi(x) = \log_+ \frac{h(x)}{\lambda}, \tag{17}$$

where $\lambda = \lambda(r) > 0$ is an appropriate parameter depending on $r > 0$. When the optimal function ξ is as in (17), 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

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

Let $\lambda > 0$ and denote by $\tilde\Pi$ the σ-finite measure with $\frac{d\tilde\Pi}{d\Pi}(x) = h(x)$. Now observe that for a non-negative function ξ satisfying the constraint (16) one has

$$\int h(x)\exp(-\xi(x))d\Pi(x) \geqslant \int \Big[\zeta(\xi(x)) + \lambda\frac{\xi(x)}{h(x)}\Big] d\tilde\Pi(x) - \lambda r \tag{18}$$

$$\geqslant \int \bar\zeta\left(\frac{\lambda}{h(x)}\right) d\tilde\Pi(x) - \lambda r. \tag{19}$$

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

Next, we note that one has equality in (18) iff

$$\begin{cases} \int \xi(x)\, d\Pi(x) = r \text{ and} \\ \xi(x) = 0 \text{ for } \Pi\text{-a.e. } x \text{ with } h(x) = 0. \end{cases} \tag{20}$$

We need to look for a non-negative function ξ and a parameter $\lambda > 0$ such that (20) is valid and such that

$$\bar{\zeta}\left(\frac{\lambda}{h(x)}\right) = \zeta(\xi(x)) + \frac{\lambda}{h(x)}\xi(x) \quad \text{for } \tilde{\Pi}\text{-a.e. } x. \tag{21}$$

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 (21) is equivalent to

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

Together with (20) a sufficient criterion for ξ 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 λ 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

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

Note that if ξ does not coincide with ξ_λ Π-a.e. (where λ is such that $g(\lambda) = r$), then one of the inequalities (18) or (19) is a strict inequality so that ξ does not minimize the target function. $\square$

Proof of Theorem 3.3. Fix $\varepsilon > 0$. The standard addition on $\mathbb{D}[0,\infty)$ is measurable with respect to the Borel σ-field; moreover, the ordered times of the jumps bigger than ε are measurable so that we can decompose X into a pure jump process consisting of all jumps bigger than ε and a further process in a measurable way. Due to Proposition 3.2 we can thus assume without loss of generality that X is a pure jump process with jumps bigger than ε. 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 prove that for an arbitrarily fixed reconstruction $\hat{X}$ with $I(X;\hat{X}) \leqslant r$ one has

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

where $c>0$ is a universal constant.

We let

$$\pi : L_1[0,1] \to \ell_1^n, \qquad (x_t) \mapsto \Big(\Big| \int_{il}^{(i+1)l} (2\mathbb{1}_{\{t \geqslant (2i+1)l/2\}} - 1) x_t \, \frac{dt}{l} \Big| \Big)_{i=0,\dots,n-1}$$

and consider

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

The map π is l^{-1}-Lipschitz continuous so that

$$\mathbb{E}[\|Y - \hat{Y}\|_{\ell_1^n}] \leqslant l^{-1} \mathbb{E}[\|X - \hat{X}\|]. \tag{22}$$

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

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

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

$$\mathbb{E}[\|Y - \hat{Y}\|_{\ell_1^n}] \geqslant \mathbb{E}\Big[\sum_{i=0}^{n-1} \mathbb{E}\big[|\mathbb{1}_{A_i} Y_i - \mathbb{1}_{A_i} \hat{Y}_i| \big| H, Y^{i-1} \big] \Big]. \tag{23}$$

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 $[\mathbb{1}_{\{H_i<0\}} H_i/2, \mathbb{1}_{\{H_i>0\}} H_i/2]$, cf. e.g. Propoposition 13.15 in [3]. Therefore,

$$\mathbb{E}\big[|\mathbb{1}_{A_i} Y_i - \mathbb{1}_{A_i} \hat{Y}_i| \; \big| H, Y^{i-1} \big] \geqslant D(R_i | \mathcal{U}[0, |H_i|/2], |\cdot|), \tag{24}$$

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

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

Together with (23) and (24) we arrive at

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

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

$$\mathbb{E}[\|Y-\hat{Y}\|_{\ell_1^n}] \geqslant c\int |H_i|\,e^{-R_i}\,d\Pi(\omega,i). \tag{25}$$

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

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

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 $\int R_i\,d\Pi(\omega,i)\leqslant r$. 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 (22) and (25) we get that

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

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 $F_1(\varepsilon)\geqslant 18$, one has*

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

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

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

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 width 2ε. Assuming that the process is symmetric the process leaves the strip 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 ε one needs to describe at least in which direction the process left the strip 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 $\bar{X}$ denote an independent copy of X and observe that for $s \in (0, 1]$

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

The process $X - \bar{X}$ is a symmetric Lévy process and the functions describing its complexity are

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

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

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

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

Then

$$\mathbb{P}(T \geqslant t) \leqslant \frac{9}{4F_1(2\varepsilon)\, t}.$$

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

$$T^* = \inf\{t \geqslant 0 : |X^*_t| \geqslant \varepsilon\}$$

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

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

Therefore,

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

Consequently,

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

and the assertion follows immediately. □

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

$$D(d\log 2d + (1-d)\log 2(1-d) \mid \mathbb{P}_Y, \rho_{\mathrm{Ham}}) \geqslant d,$$

where ρ_{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 ρ 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 $d \in [0,1/2]$:

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

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

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

□

The proof of the lower bound in Theorem 3.5 is based on a comparison with a simpler distortion rate function. For $q \in [0,1/2]$ let μ_q denote the measure that assigns probabilities q to ±1 and $1-2q$ to 0. Moreover denote by $\mu_q^{\otimes n}$ its product measure, consider the distortion measure

$$\rho(x,\hat{x}) = \mathbb{1}_{\{x\cdot\hat{x}=-1\}} \qquad (x \in \{\pm1, 0\}, \hat{x} \in \{\pm1\})$$

and denote

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

As reconstruction we allow any $\{\pm1\}^n$-valued random vector.

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

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

where

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

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

$$\pi : L_1[0,1] \to \ell_1^n, \qquad (x_t) \mapsto \Big(\Big|\int_{il}^{(i+1)l} (2\mathbb{1}_{\{t\,\geqslant\,(2i+1)l/2\}} - 1)x_t\,\frac{dt}{l}\Big|\Big)_{i=0,\dots,n-1}.$$

The map π 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

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

Recalling the Lipschitz continuity of π we get that

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

Therefore,

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

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

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

Next, we show that $D(r|\mu_q^{\otimes n}, \rho_n, s)$ is increasing in q. Indeed, let $0 \leqslant q < q' \leqslant 1/2$, let Z denote an $\mu_{q'}^{\otimes n}$ distributed r.v., and let $\hat{Z}$ denote a reconstruction for Z with $I(Z; \hat{Z}) \leqslant r$. 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

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

It remains to prove that $\mathbb{P}(Y_i \geqslant \varepsilon/4) \geqslant \frac{1}{8}\big(1 - \frac{9}{F_1(2\varepsilon)\, l}\big)$. 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 ν. Denote $T^+ = \inf\{t \geqslant 0 : \tilde{X}_t \geqslant \varepsilon \text{ or } t \geqslant l/2\}$ and observe that

$$\begin{aligned}
\mathbb{P}\big(Y_i \geqslant \frac{\varepsilon}{4}\big) &\geqslant \mathbb{P}\Big(-\int_0^{l/2} \tilde{X}_{-t}\, dt \geqslant 0, T \leqslant l/4, \int_T^{l/2} [\tilde{X}_t - \tilde{X}_T]\, dt \geqslant 0\Big) \\
&= \mathbb{P}\Big(\int_0^{l/2} \tilde{X}_{-t}\, dt \leqslant 0\Big)\, \mathbb{P}(T \leqslant l/4)\, \mathbb{P}\Big(\int_T^{l/2} [\tilde{X}_t - \tilde{X}_T]\, dt \geqslant 0 | T \leqslant l/4\Big) \\
&= \frac{1}{4}\, \mathbb{P}(T^+ \leqslant l/4).
\end{aligned}$$

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

$$\mathbb{P}(T^+ \leqslant l/4) \geqslant \frac{1}{2} \mathbb{P}(T \leqslant l/4) \geqslant \frac{1}{2}\Big(1 - \frac{9}{F_1(2\varepsilon)\, l}\Big)$$

so that

$$\mathbb{P}\big(Y_i \geqslant \frac{\varepsilon}{4}\big) \geqslant \frac{1}{8}\Big(1 - \frac{9}{F_1(2\varepsilon)\, l}\Big).$$

□

Lemma 3.10. *Let μ^{Ber} and ρ_{Ham} denote the Bernoulli distribution and the Hamming distance, respectively. Then*

$$D(r|\mu_q,\rho) \geqslant 2q\, D\Big(\frac{r}{2q}\Big|\mu^{\mathrm{Ber}},\rho_{\mathrm{Ham}}\Big).$$

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

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

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

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

Next, we write

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

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

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

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

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

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

Additionally, there exists a universal constant $C_3>0$ such that $n\geqslant C_3F_1(2\varepsilon)$. Next, we shall apply Proposition 3.9. We fix $r_0<\log 2$ arbitrarily and set $r=\frac{1}{8}nr_0$. Then $r\geqslant C_1\,F_1(2\varepsilon)$ for some constant C_1 only depending on the choice of r_0. Thus with Proposition 3.9 one gets

$$D(C_1F_1(2\varepsilon),s) \geqslant D(r,s) \geqslant \frac{\varepsilon}{4n} D\Big(\frac{1}{8}nr_0\Big|\mu_q^{\otimes n},\rho_n,s\Big). \tag{26}$$

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

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

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

$$D\Big(\frac{1}{8}nr_0\Big|\mu_q^{\otimes n},\rho_n\Big) \geqslant n\,2q\,D\Big(\frac{1}{16q}r_0\Big|\mu^{\mathrm{Ber}},\rho_{\mathrm{Ham}}\Big).$$

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

$$D(C_1F_1(2\varepsilon),1) \geqslant \frac{\varepsilon}{2}qC_4 \geqslant C_2\,2\varepsilon,$$

where $C_2=C_4/64$. Switching from 2ε to ε finishes the proof of the first assertion. $\square$

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

Lemma 3.11. *Let $\rho : \mathbb{R} \times \mathbb{R} \to [0, \infty]$ 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*

$$\mathbb{E}[\rho(U_1, u^*)^2] < \infty, \tag{27}$$

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

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

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

As one can see in the proof the moment condition (27) can be easily relaxed. Similar ideas are used in [8] 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

- $\{(u^*, \dots, u^*)\} \subset \mathcal{C}(n) \subset \mathbb{R}^n$,
- $\log |\mathcal{C}(n)| \lesssim nr$,
- $\lim_{n \to \infty} \mathbb{P}(\mathcal{T}(n)) = 1$ 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 $I(U^{(n)}, \hat{U}^{(n,1)}) \leqslant nr$, 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

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

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

Next, we use that

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

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

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

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

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

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 $\log \frac{d\mathbb{P}_{U^{(n)},\hat U^{(n)},J}}{d\mathbb{P}_{U^{(n)}}\otimes \mathbb{P}_{\hat U^{(n,1)}}\otimes\delta_1} \leqslant nr$ on $\{J=1\}$. Moreover, one has $\log \frac{d\mathbb{P}_{U^{(n)},\hat U^{(n)},J}}{d\mathbb{P}_{U^{(n)}}\otimes Q^*\otimes\delta_2} \leqslant \log|\mathcal{C}(n)|$ on $\{J=2\}$. Consequently, we can continue with

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

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

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

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

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

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

Proof of Theorem 3.5, 2nd 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 ∞ as ε tends to 0.

We recall estimate (26):

$$D(C_1F_1(2\varepsilon),s) \geqslant D(r,s) \geqslant \frac{\varepsilon}{4n} D\Big(\frac{1}{8}nr_0\Big|\mu_q^{\otimes n},\rho_n,s\Big).$$

Now we conclude with Lemma 3.11 that

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

The assertion follows along the lines of the proof of the first statement. □

Acknowledgements

The authors would like to thank the anonymous referees for reading the manuscript carefully and for their suggestions. The research of the first-mentioned author was supported by the DFG Research Center "Matheon – Mathematics for key technologies" in Berlin.

References

[1] F. Aurzada, S. Dereich, M. Scheutzow, and C. Vormoor. High resolution quantization and entropy coding of jump processes. Preprint, URL: http://arxiv.org/abs/0712.0964, 2007.

[2] J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge: Cambridge University Press, 1996.

[3] L. Breiman, *Probability*, volume 7 of *Classics in Applied Mathematics.* Philadelphia, USA: SIAM, 1993.

[4] T. M. Cover and J. A. Thomas. *Elements of information theory.* Hoboken, NJ: Wiley, 2. ed., 2006.

[5] J. Creutzig. Approximation of SαS Lévy processes in L_p norm. *J. Approx. Theory* 140(1):71–85, 2006.

[6] S. Dereich. High resolution coding of stochastic processes and small ball probabilities. Ph.D. Dissertation, TU Berlin,
URL: http://edocs.tu-berlin.de/diss/2003/dereich_steffen.htm, 2003.

[7] S. Dereich. The coding complexity of diffusion processes under supremum norm distortion. *Stochastic Process. Appl.* 118(6):917–937, 2008.

[8] S. Dereich. The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion. *Stochastic Process. Appl.* 118(6):938–951, 2008.

[9] S. Dereich, J. Creutzig, T. Müller-Gronbach, and K. Ritter. Infinite-dimensional quadrature and approximation of distributions. to appear in: *Found. Comput. Math.*

[10] S. Dereich, F. Fehringer, A. Matoussi, and M. Scheutzow. On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *J. Theoret. Probab.* 16(1):249–265, 2003.

[11] S. Dereich and M. Scheutzow. High resolution quantization and entropy coding for fractional brownian motion. *Electron. J. Probab.* 11:700–722, 2006.

[12] P. Elias. Universal codeword sets and representations of the integers. *IEEE Trans. Inform. Theory* 21:194–203, 1975.

[13] S. Ihara. *Information theory for continuous systems.* Singapore: World Scientific, 1993.

[14] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Internat. J. Comput. Math.* 2:157–168, 1968.

[15] H. Luschgy and G. Pagès. Sharp asymptotics of the functional quantization problem for Gaussian processes. *Ann. Probab.* 32(2):1574–1599, 2004.

[16] H. Luschgy and G. Pagès. Functional quantization of a class of Brownian diffusions: a constructive approach. *Stochastic Process. Appl.* 116(2):310–336, 2006.

[17] H. Luschgy and G. Pagès. Functional quantization rate and mean pathwise regularity of processes with an application to Lévy processes. *Ann. Appl. Probab.* 18(2):427–469, 2008.

[18] G. Pagès and J. Printems. Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods Appl.* 11(4):407–446, 2005.

[19] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York).* Berlin: Springer-Verlag, second edition, 2004.

[20] K. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 in *Cambridge Studies in Advanced Mathematics*, Cambridge: Cambridge University Press, 1999.

[21] C. Vormoor. *High resolution coding of point processes and the Boolean model*. PhD thesis, Technische Universität Berlin,
URL: http://opus.kobv.de/tuberlin/volltexte/2007/1502/, 2007.