

High resolution quantization and entropy coding of jump processes

Frank Aurzada, Steffen Dereich, Michael Scheutzow, and
Christian Vormoor

*Technische Universität Berlin, Institut für Mathematik, Sekr. MA 7-5, Straße des
17. Juni 136, 10623 Berlin, Germany.*

Abstract

We study the quantization problem for certain types of jump processes. The probabilities for the number of jumps are assumed to be bounded by Poisson weights. Otherwise, jump positions and increments can be rather generally distributed and correlated. We show in particular that in many cases entropy coding error and quantization error have distinct rates. Finally, we investigate the quantization problem for the special case of \mathbb{R}^d -valued compound Poisson processes.

Key words: High resolution quantization; entropy coding; complexity; jump process; compound Poisson process; Lévy process; metric entropy

1 Introduction and results

1.1 Statement of the problem

In this article, we study the quantization- and entropy coding problem for certain types of jump processes. Given a random variable X , the aim is to find a good approximation \hat{X} to X that satisfies a particular *complexity constraint*.

Let $s > 0$, X be a random variable in a measurable space (E, \mathcal{E}) , ρ a distortion measure on E (i.e. a measurable, symmetric function $\rho : E \times E \rightarrow \mathbb{R}_{\geq 0}$ with $\rho(x, y) = 0$ iff $x = y$),

Email addresses: aurzada@math.tu-berlin.de (Frank Aurzada),
dereich@math.tu-berlin.de (Steffen Dereich), ms@math.tu-berlin.de (Michael Scheutzow),
christian.vormoor@devk.de (Christian Vormoor).

and $r \geq 0$. Then we define the *quantization error* as follows:

$$D^{(q)}(r | X, \rho, s) := \inf \left\{ \left(\mathbb{E} \min_{a \in \mathcal{C}} \rho(X, a)^s \right)^{1/s} : \log \#\mathcal{C} \leq r \right\}.$$

The number $D^{(q)}$ represents the best-achievable average error when encoding the signal X with r nats. The term ‘nats’ is used instead of ‘bits’, since we calculate the amount of information using the natural logarithm. Further, we investigate the *entropy coding error*, which can be understood as the average error when encoding the signal X using – on average – r nats:

$$D^{(e)}(r | X, \rho, s) := \inf \left\{ \left(\mathbb{E} \rho(X, \hat{X})^s \right)^{1/s} : \hat{X} \text{ random var. with } H(\hat{X}) \leq r \right\},$$

where H is the (discrete) entropy of a random variable:

$$H(X) := \begin{cases} -\sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x) & X \text{ discrete,} \\ \infty & \text{otherwise.} \end{cases}$$

In slight misuse of notation we also write $D^{(q)}(r | X, \|\cdot\|, s)$ if $\rho(x, y) = \|x - y\|$ for a norm distortion $\|\cdot\|$. Analogously, we deal with the entropy coding error. We recall that $D^{(e)}(r | X, \rho, s) \leq D^{(q)}(r | X, \rho, s)$.

The problems described above arise naturally in coding theory, where for instance, the complexity of a signal has to be reduced due to capacity restrictions of a channel or simply (lossy) data compression is considered (see for instance [6] for a general account on coding theory and [15] for a historic outline of the information constraints). Beyond these information-theoretic applications, the quantization error is tightly related to certain quadrature problems: the quantization error can be defined equivalently as the worst-case error of a particular quadrature problem. Moreover, further quadrature problems are linked to the quantization problem via estimates involving both quantities. Recent results in that direction can be found in [7] (see also [19] for earlier results).

The analysis of the quantization- and entropy coding error started in the 40s of the 20th century. At that time research was mainly focused on finite-dimensional signals; and the numerous publications mainly appeared in the engineering literature. A mathematical account of the results for finite-dimensional signals is provided by [13]. Since about 2000 researchers are attracted by the problem in the case where the original signal is infinite-dimensional. A series of articles followed on (infinite-dimensional) random vectors X that are Gaussian (see for instance [11], [16], [12]), diffusions ([17], [9], [10]), and Lévy processes ([18], [2]).

In this article, we provide asymptotic estimates for the quantization- and entropy coding error for certain jump processes. The results are shown to be sharp in several cases.

In contrast to the (infinite-dimensional) settings studied before, there is a qualitative difference in the (best-achievable) approximation error induced by the two constraints.

1.2 Some notation and the model

Let us now introduce the jump processes that we investigate in this article.

We define the space $\mathbf{D}([0, 1[, E)$ to be the space of all functions $f : [0, 1[\rightarrow E$ that are piecewise constant and possess a finite number of jumps, where if f has a jump at t from the value $a \in E$ to $b \in E$, then $f(t) = b$. We endow $\mathbf{D}([0, 1[, E)$ with the σ -field induced by the projections.

In the sequel, $X = (X(t))_{t \in [0, 1[}$ denotes a $\mathbf{D}([0, 1[, E)$ -valued random vector. We denote by N_X the random number of jumps of X , let $0 < Y_1 < \dots < Y_{N_X} < 1$ be the *jump positions* of X , and set $Y_0 = 0$ and $Y_{N_X+1} = 1$. Moreover, we denote by

$$Z_i := \rho(X(Y_{i-1}), X(Y_i))$$

the *moduli of the increments* and, in the case where E is a linear space, we denote by

$$Z^{(i)} := X(Y_i) - X(Y_{i-1})$$

the *increments*.

As distortion measure on $\mathbf{D}([0, 1[, E)$ we consider

$$\rho_{\mathbf{D}}(f, g) := \int_0^1 \rho(f(t), g(t)) dt, \quad f, g \in \mathbf{D}([0, 1[, E), \quad (1)$$

where ρ is a distortion measure on E . It is straightforward to extend the results of this paper to the distortion measure $\rho_{\mathbf{D}}^p(f, g) = (\int_0^1 \rho(f(t), g(t))^p dt)^{1/p}$, with $1 \leq p < \infty$.

Our lower bounds require that the jump positions constitute a *Poisson point process* with intensity λ . The upper bounds on the complexity are proven under weaker assumptions on X . Here, we only assume that the total number of jumps can be estimated against the probability weights of a Poisson random variable:

$$\mathbb{P}(N_X = k) \leq \frac{\lambda^k}{k!} e^{-\lambda} K, \quad k \geq 0, \quad (2)$$

where $\lambda > 0$ and $K \geq 1$ are some fixed parameters. In particular, one can choose $K = 1$, if the jump positions are induced by a Poisson point process.

Sometimes we shall also impose the following condition:

(*) The jump positions are independent of the jump destinations, which means that, given the event $\{N_X = k\}$, the vector (Y_1, \dots, Y_k) is independent of the vector $(X(Y_0), \dots, X(Y_k))$.

Let us introduce some more notation. Firstly, we make use of the concept of metric entropy. If ρ is a distortion measure on E we define its covering numbers by

$$N(E, \rho, \varepsilon) := \min\{n \in \mathbb{N} : \exists x_1, \dots, x_n \in E \forall x \in E \exists i : \rho(x, x_i) \leq \varepsilon\}.$$

A set $\{x_1, \dots, x_n\}$ for which the defining property of N holds is called an ε -net of E . Note that in general one has to assume that $N(E, \rho, \varepsilon)$ is well-defined, i.e. that for all $\varepsilon > 0$ there is an ε -net of E . This is ensured if, for example, (E, ρ) is a precompact metric space. We also introduce the inverse concept of $D^{(q)}$, which we call $d^{(q)}$, given by

$$d^{(q)}(\varepsilon | X, \rho, s) := \inf\{n \geq 1, n \in \mathbb{N} : D^{(q)}(\log n | X, \rho, s) \leq \varepsilon\}.$$

In other words, $d^{(q)}$ is the number of points needed to quantize with error at most ε , i.e. roughly it is the inverse function of $D^{(q)}(\log(\cdot))$.

We shall also need the notation of strong and weak asymptotics. Namely, we write $f \lesssim g$, if $\limsup f/g \leq 1$. Analogously, $f \gtrsim g$ is defined. Furthermore, $f \sim g$ means $\lim f/g = 1$. We also use $f \approx g$ if $0 < \liminf f/g \leq \limsup f/g < \infty$. Finally, throughout the article λ_d denotes the d -dimensional Lebesgue measure.

The paper is organized as follows. In the rest of this section we state the main results. In Section 1.3, we state the upper bounds for both quantities under various additional assumptions. In Section 1.4, the upper bounds are complemented by corresponding lower bounds. In particular, we obtain that the upper and lower bounds are tight in many cases. Finally, Section 1.5 is devoted to the particular setting where X is a compound Poisson process. The proofs for the upper bounds can be found in Section 2. There, explicit coding strategies are constructed. The lower bounds are proven in Sections 3 and 4 for quantization- and entropy coding, respectively. The proofs for the lower bounds of the quantization error rely on a small ball argument, whereas the lower bounds for the entropy coding error are derived using the Shannon lower bound for a related problem.

1.3 Upper bounds

Our first result concerns the case where the space E has finite covering numbers. In the case where (E, ρ) is a metric space, this corresponds to the assumption that E is precompact.

Theorem 1 Assume that $w := \sup_{x,y \in E} \rho(x,y) < \infty$ and that the upper box dimension

$$\gamma := \limsup_{\varepsilon \rightarrow 0} \frac{\log N(E, \rho, \varepsilon)}{\log 1/\varepsilon} \quad (3)$$

is finite. Then

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \gtrsim \sqrt{\frac{2}{s(1+\gamma)}} r \log r. \quad (4)$$

Theorem 2 Assume that $N(E, \rho, \varepsilon) < \infty$ for all $\varepsilon > 0$ and that $w := \sup_{x,y \in E} \rho(x,y) < \infty$.

(a) For all $r > r_0 = r_0(\lambda)$,

$$D^{(e)}\left(K\left(\lambda r + (\lambda + 1) \log N(E, \rho, e^{-r})\right) \mid X, \rho_{\mathbf{D}}, s\right) \leq C_s(w + 1)K^{1/s}e^{-r}, \quad (5)$$

where the constant C_s depends on s only, and K and λ are the constants from (2).

(b) In particular, if the jump positions are distributed according to a Poisson point process with rate λ , we obtain for $r > r_0 = r_0(\lambda)$

$$D^{(e)}(\lambda r + (\lambda + 1) \log N(E, \rho, e^{-r}) \mid X, \rho_{\mathbf{D}}, s) \leq C_s(w + 1)e^{-r}. \quad (6)$$

(c) In the case of a discrete space $E = \{x_1, \dots, x_q\}$ we even have the more precise estimate

$$D^{(e)}(K(\lambda r + (\lambda + 1) \log q) \mid X, \rho_{\mathbf{D}}, s) \leq 4K^{1/s}w \min(1, \lambda)^{1/s} e^{-r}, \quad (7)$$

for $r > r_0 = r_0(\lambda)$.

Theorem 2 can be interpreted in the following way. In order to quantize with error e^{-r} one needs, on average, λr nats to encode the jump positions, $\lambda \log N(E, \rho, e^{-r})$ nats in order to encode the increments, and another $\log N(E, \rho, e^{-r})$ nats in order to encode the initial position $X(0)$. In particular, the same result can be proved without the $\log N(E, \rho, e^{-r})$ term if the initial value of the process is deterministic.

Let us compare Theorem 1 and Theorem 2 in the case where $N(E, \rho, \varepsilon) \leq q\varepsilon^{-\gamma}$. We point out that the asserted rate of the quantization error is different to the one of the entropy coding error. As we will see below neither the quantization error bounds nor the entropy coding error bounds can be improved significantly.

Finally note that $N(E, \rho, \varepsilon) < \infty$ for all $\varepsilon > 0$ does not necessarily imply that $w := \sup_{x,y \in E} \rho(x,y) < \infty$ if ρ does not satisfy the triangle inequality.

For the remainder of this subsection, let us assume that $(E, \|\cdot\|)$ is a normed linear space with distortion measure $\rho(x,y) = \|x - y\|$. We assume that the jump destinations of X

(and thus the increments $Z^{(i)}$) are independent of the jump positions (condition (*)). Furthermore, assume that the increments, conditioned upon $N_X = k$, are identically distributed (not necessarily independent among each other) with the same law as the E -valued random variable, say, $Z^{(1)}$. Furthermore we assume that $X(0)$ is deterministic, i.e. that for some $x_0 \in E$ $X(0) = x_0$ a.s.

Theorem 3 *Under the above assumptions the following statements are true.*

(a) *If*

$$\gamma := \limsup_{\varepsilon \rightarrow 0} \frac{\log d^{(q)}(\varepsilon | Z^{(1)}, \|\cdot\|, s)}{\log 1/\varepsilon} \in [0, \infty[,$$

then (4) is valid for the newly defined γ .

(b) *If $d^{(q)}(\varepsilon | Z^{(1)}, \|\cdot\|, s) < \infty$ for all $\varepsilon > 0$, then*

$$D^{(e)}\left(K\left(\lambda r + \lambda \log d^{(q)}(e^{-r} | Z^{(1)}, \|\cdot\|, s)\right) \mid X, \rho_{\mathbf{D}}, s\right) \leq C e^{-r}.$$

holds with some constant $C > 0$ depending on the parameters s, K, λ , and $\mathbb{E}\|Z^{(1)}\|^s$.

Theorem 3 relates the complexity of coding X to that of coding the increments. If the assumptions of both Theorems 2 and 3 are satisfied, then the bounds of the latter theorem provide a better estimate since in general $d^{(q)}(\varepsilon) \leq N(\varepsilon)$ (see Lemma 20). However, note that in contrast to Theorems 1 and 2, Theorem 3 requires that the increments are identically distributed and independent of the jump positions. In case of Theorems 1 and 2, this is not necessary since, by assumption, the space E is sufficiently well-structured (in the sense of small metric entropy N).

Let us remark that the assumption in Theorem 3 that $X(0)$ be deterministic is for simplicity only. If instead $X(0)$ is a random variable in E , one has to add $d^{(q)}(\varepsilon | X(0), \rho, s)$ to the average number of nats needed to encode X conditioned upon $X(0)$.

Finally we mention that one can also prove counterparts to assertions (b) and (c) of Theorem 2 in the setting of Theorem 3.

1.4 Lower bounds

As an illustration consider the case of a discrete space E , namely let $E = \{x_1, \dots, x_q\}$, which was first studied in [20]. Then $N(E, \rho, \varepsilon) \leq q$, and we thus obtain from Theorems 1 and 2:

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \gtrsim \sqrt{\frac{2}{s} r \log r} \quad \text{and} \quad -\log D^{(e)}(Kr \mid X, \rho_{\mathbf{D}}, s) \gtrsim \frac{r}{\lambda}. \quad (8)$$

Now we ask for lower bounds. Clearly, one cannot expect a non-trivial lower bound when only assuming (2). Thus, let us assume in this subsection that the jump positions constitute a Poisson point process and that condition (*) holds. In this case, we show in Theorem 4 that the order of $D^{(q)}$ in (8) is in fact the true order on this scale. Below, in Theorem 7, we show that the order of $D^{(e)}$ is the correct one, too.

We consider a more general situation than a finite, discrete space. We only have to assume that there is sufficient uncertainty in the model in order to ensure that every jump indeed has to be encoded.

Concretely, assume that condition (*) holds and that the jump positions form a Poisson point process. Furthermore, we assume that, given the event that k jumps occur ($\{N_X = k\}$), the moduli of the increments Z_1, \dots, Z_k are such that there are $\varepsilon_0 > 0$ and $\delta_0 > 0$ (independent of k) such that for all $i = 1, \dots, k$, $\mathbb{P}(Z_i > \varepsilon_0 \mid N_X = k) \geq \delta_0$. Additionally, we now impose that (E, ρ) is a *metric* space.

Theorem 4 *Under the above assumptions,*

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \lesssim \sqrt{\frac{2}{s} r \log r}.$$

In particular, for a discrete metric space $E = \{x_1, \dots, x_q\}$,

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \sim \sqrt{\frac{2}{s} r \log r}.$$

Note that in view of (7) the rates for quantization error and entropy coding error must be different in case of a discrete metric space $E = \{x_1, \dots, x_q\}$. Moreover, the order of convergence of the quantization error depends strongly on the moment s . In particular, one has for two distinct moments $0 < s < s'$ that

$$\lim_{r \rightarrow \infty} \frac{D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s)}{D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s')} = 0.$$

This contrasts earlier results on quantization where the same order of convergence is obtained for all moments $s > 0$.

Let us consider a simple example.

Example 5 *Let X be an alternating Poisson process, i.e.*

$$X(t) = \sum_{i=1}^{N(t)} (-1)^{i-1}, \tag{9}$$

where $(N(t))_{t \in [0,1]}$ is a Poisson (counting) process with rate λ (cf. Section 1.5) with the natural metric $|\cdot|$ on $E = \{0, 1\}$. Then Theorem 4 yields

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \sim \sqrt{\frac{2}{s} r \log r}.$$

Remark 6 Recall that the assertions of the upper bounds, do not require that (E, ρ) is a metric space. The statement is valid for any distortion measure ρ . However, the lower bound from Theorem 4 fails for general distortion measures. This can be seen from the following simple example. Let $E = \{0, 1\} \cup \{2^{-n}, n \geq 1\}$ and $\rho(0, 1) = 1$, $\rho(0, 2^{-n}) = \rho(1, 2^{-n}) = 2^{-n}$, $\rho(2^{-n}, 2^{-m}) = 1$ for $n \neq m$ and $n, m \geq 1$. Note that this is not a metric space.

Consider the alternating Poisson process, i.e. the model from (9). Then X satisfies the assumptions of Theorem 4, in particular those for the moduli of the increments (since $Z_i = 1$), but $D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) = 0$, for all $r \geq 0$.

Next, we will prove a lower bound for the entropy coding error.

Theorem 7 Let X be a jump process satisfying condition (*). We assume that the jump positions (Y_i) form a Poisson point process with rate λ . Moreover, we suppose that ρ defines a metric on E and that a.s. the moduli of the jumps are bounded from below by $\varepsilon_0 > 0$. Then for all $s \geq 1$ and all sufficiently large r

$$D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \geq \varepsilon_0 C \min(1, \lambda) e^{-r/\lambda},$$

where $C > 0$ is an absolute constant.

Remark 8 The lower bound in Theorem 7 is actually shown to hold for the distortion rate function $D(r \mid X, \rho, s)$ defined in Section 4.

We obtain the following corollary as a special case.

Corollary 9 Let X satisfy the conditions of Theorem 7. Assume additionally that $X(0)$ is deterministic and consider the case of a discrete metric space $E = \{x_1, \dots, x_q\}$ with $w := \max_{x, y \in E} \rho(x, y)$. Then for $s \geq 1$

$$C_1 \varepsilon_0 \min(1, \lambda) e^{-r/\lambda} \leq D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \leq C_2 q w \min(1, \lambda)^{1/s} e^{-r/\lambda},$$

for large enough r and absolute constants $C_1, C_2 > 0$.

The corollary follows immediately from part (c) of Theorem 2 and the remark after it and Theorem 7. This result shows that the bounds for the entropy coding error in Theorems 2 and 3 are tight.

Example 10 Consider again the alternating Poisson process from (9) with the natural metric $|\cdot|$. Then Corollary 9 yields, for all $s \geq 1$,

$$C_1 \min(1, \lambda) e^{-r/\lambda} \leq D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \leq C_2 \min(1, \lambda)^{1/s} e^{-r/\lambda},$$

for large enough r and absolute constants $C_1, C_2 > 0$.

Example 11 Let us illustrate the influence of a random initial position on our estimates. For this purpose, consider an alternating Poisson process with random initial position, i.e.

$$X(t) = X(0) + \sum_{i=1}^{N(t)} (-1)^{i-1+X(0)},$$

where $X(0)$ equals 0 and 1 with probability $1/2$, respectively, cf. [20]. Our Theorem 2, part (c), and Theorem 7 yield

$$C_1 \min(1, \lambda) e^{-r/\lambda} \leq D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \leq C_2 \min(1, \lambda)^{1/s} 2^{1/\lambda} e^{-r/\lambda},$$

for all $s \geq 1$ and all large enough r and absolute constants $C_1, C_2 > 0$.

1.5 Application to compound Poisson processes in \mathbb{R}^d

As an application of our results, let us determine the coding complexity of \mathbb{R}^d -valued compound Poisson processes. Recall that a Lévy process with finite Lévy measure is a compound Poisson process with the following structure, cf. e.g. [3].

Let $(N(t))_{t \geq 0}$ be a Poisson (counting) process with intensity $\lambda > 0$, i.e. let $N(t) := \max\{n \geq 0 : \sum_{j=1}^n e_j \leq \lambda t\}$, where (e_j) are i.i.d. standard exponential random variables. Consider

$$X(t) = \sum_{i=1}^{N(t)} Z^{(i)}, \quad t \in [0, 1[, \quad (10)$$

where the $Z^{(i)}$, $i = 1, 2, \dots$, are i.i.d. and distributed according to any probability distribution in \mathbb{R}^d with $\mathbb{P}(Z^{(1)} = 0) = 0$. Note that this notation is consistent with the one employed above for the increments. Note furthermore that for compound Poisson processes condition (*) is satisfied.

We consider the distortion measure

$$\|X\|_1 := \int_0^1 \|X(t)\|_\infty dt,$$

which of course coincides with $\rho_{\mathbf{D}}$ for $\rho = \|\cdot\|_\infty$, where as usual $\|x\|_\infty := \max_{i=1, \dots, d} |x_i|$. However, one can replace $\|\cdot\|_\infty$ by any norm on \mathbb{R}^d , which would change only the constants.

Theorem 3 yields the following corollary.

Corollary 12 *Let X be a compound Poisson process as defined in (10) and $s > 0$.*

(a) *Assume that*

$$\gamma := \limsup_{\varepsilon \rightarrow 0} \frac{\log d^{(q)}(\varepsilon | Z^{(1)}, \|\cdot\|_\infty, s)}{\log 1/\varepsilon} \quad (11)$$

is finite. Then

$$-\log D^{(q)}(r | X, \rho_{\mathbf{D}}, s) \gtrsim \sqrt{\frac{2}{s(1+\gamma)}} r \log r.$$

(b) *Let $d^{(q)}(\varepsilon | Z^{(1)}, \rho, s) < \infty$ for all $\varepsilon > 0$. Then, for $r \geq r_0$ and a constant $C = C(s, \lambda, \mathbb{E}\|Z^{(1)}\|_\infty^s)$, we have*

$$D^{(e)}(\lambda r + (\lambda + 1)d^{(q)}(e^{-r} | Z^{(1)}, \rho, s) | X, \rho_{\mathbf{D}}, s) \leq C e^{-r}.$$

Alternatively, one can study the consequences of Theorems 1 and 2 if one has additional information on the range of X .

As for lower bounds we can apply Theorem 4, which gives the following.

Corollary 13 *Let X be a compound Poisson process as defined in (10) and $s > 0$. Then*

$$-\log D^{(q)}(r | X, \rho_{\mathbf{D}}, s) \lesssim \sqrt{\frac{2}{s}} r \log r.$$

If additionally (11) holds with $\gamma = 0$, then

$$-\log D^{(q)}(r | X, \rho_{\mathbf{D}}, s) \sim \sqrt{\frac{2}{s}} r \log r.$$

We obtain a similar result in the case that the distribution of the increments has an absolutely continuous component.

Theorem 14 *Let X be a compound Poisson process as defined in (10) and $s > 0$. Assume that the distribution of $Z^{(1)}$ has an absolutely continuous component. Then*

$$-\log D^{(q)}(r | X, \rho_{\mathbf{D}}, s) \lesssim \sqrt{\frac{2}{s(1+d)}} r \log r.$$

If additionally (11) holds with $\gamma = d$ then

$$-\log D^{(q)}(r | X, \rho_{\mathbf{D}}, s) \sim \sqrt{\frac{2}{s(1+d)}} r \log r.$$

Theorems 4 and 14 show that the upper bound for the quantization rate in Theorems 1 and 3 (and thus Corollary 12) cannot be improved in general (for all $\gamma \in \mathbb{N}$).

Let us finally list a corollary of Theorem 7.

Corollary 15 *Let X be a compound Poisson process as defined in (10). Assume that $\|Z^{(1)}\|_\infty > \varepsilon_0$ a.s. Then, for all $s \geq 1$,*

$$D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \geq \varepsilon_0 C \min(1, \lambda) e^{-r/\lambda},$$

for $r > r_0$ and $C > 0$ an absolute constant.

The most instructive examples of the application of the results of this subsection are given now.

Example 16 *Consider a Poisson (counting) process with intensity λ , i.e. let $Z^{(1)} = 1$. Then*

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \sim \sqrt{\frac{2}{s} r \log r}$$

and

$$C_1 \min(1, \lambda) e^{-r/\lambda} \leq D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \leq C_2 e^{-r/\lambda},$$

for $s \geq 1$, $r > r_0$, where $C_1 > 0$ is an absolute constant, and $C_2 > 0$ depends on s and λ .

Example 17 *Let $Z^{(1)}$ be uniformly distributed in $[0, 1]^d$. Then*

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \sim \sqrt{\frac{2}{s(1+d)} r \log r}$$

and

$$D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \leq C_2 e^{-r/((1+d)\lambda)},$$

for $s \geq 1$, $r > r_0$, where $C_2 > 0$ depends on s and λ . We conjecture that the order on the right-hand side is the correct one.

Example 18 *Let $Z^{(1)}$ be uniformly distributed in C , where C is the Cantor set in $[0, 1]$. Set $\gamma := \log 2 / \log 3$. Then*

$$\sqrt{\frac{1}{s} r \log r} \gtrsim -\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \gtrsim \sqrt{\frac{2}{s(1+\gamma)} r \log r}$$

and

$$D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \leq C_2 e^{-r/((1+\gamma)\lambda)},$$

for $s \geq 1, s r > r_0$, where $C_2 > 0$ depends on s and λ . We conjecture that the orders on the right-hand side, respectively, are the correct ones.

The theorems and examples presented in this subsection complement results from [2], where general real-valued Lévy processes are studied. The main result for compound Poisson processes in that paper states that, for any compound Poisson process with $\mathbb{E} \log \max(|Z^{(1)}|, 1) < \infty$ and all $s \geq 1$,

$$\log D^{(e)}(r \mid X, \rho_{\mathbf{D}}, s) \approx -r.$$

No result on the quantization error for compound Poisson processes is obtained in [2].

Our findings also improve the results in [18], where an upper bound for the quantization error of real-valued compound Poisson processes is obtained. In particular, it is shown that for the Poisson (counting) process and all $s \geq 1$,

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \gtrsim \sqrt{\frac{1}{s} r \log r}.$$

The correct rate on this scale is given in Example 16.

2 Upper bounds

In this section, we provide the proofs of the upper bounds for the quantization error and the entropy coding error stated in Theorems 1, 2, and 3, respectively. In the proofs, the following four technical lemmas are needed.

First we prove a result on the asymptotic behaviour of a certain sum occurring in the calculations.

Lemma 19 *Let $c > 0$. Then*

$$\log \left(\sum_{k=0}^{\infty} \frac{c^k}{k!} e^{-c} e^{-r/(k+1)} \right) \sim -\sqrt{2r \log r}, \quad \text{as } r \rightarrow \infty.$$

Proof: Let V be a random variable that is Poisson distributed with mean c . Then the term in question equals

$$\log \mathbb{E} e^{-r/(V+1)}.$$

By the so-called de Bruijn Tauberian theorem (cf. [4], Theorem 4.12.9), considering the Laplace transform is equivalent to considering the lower tail of $(V+1)^{-1}$. Thus, consider

$$\log \mathbb{P} \left(\frac{1}{V+1} < \varepsilon \right) = \log \mathbb{P} \left(V > \frac{1}{\varepsilon} - 1 \right) = \log \sum_{\{k > \frac{1}{\varepsilon} - 1\}} \frac{c^k}{k!} e^{-c} \sim -\frac{1}{\varepsilon} \log \frac{1}{\varepsilon},$$

where we used Stirling's Formula in the last step. Using the above-mentioned Tauberian theorem returns the asserted order of the Laplace transform, including the constant. \square

Secondly, we prove a quantization result for random variables in a space E with known metric entropy. This is needed in order to encode the increments of the process X .

Lemma 20 *Let X be any random variable on a space E . Then, for all $s > 0$ and all $\varepsilon > 0$,*

$$D^{(q)}(\log N(E, \rho, \varepsilon) \mid X, \rho, s) \leq \varepsilon.$$

In other words, $d^{(q)}(\varepsilon \mid X, \rho, s) \leq N(E, \rho, \varepsilon)$.

Proof: For given $\varepsilon > 0$ let \mathcal{C} be an ε -net of (E, ρ) . By the definition of the covering numbers, \mathcal{C} can be chosen to contain only $N(E, \rho, \varepsilon)$ elements. Thus

$$D^{(q)}(\log N(E, \rho, \varepsilon) \mid X, d, s) \leq \left(\mathbb{E} \min_{a \in \mathcal{C}} \rho(X, a)^s \right)^{1/s} \leq \varepsilon.$$

\square

Remark 21 *By using product quantization, it is clear that for a random variable X in $E^k := E \times \dots \times E$ with $\rho^k(x, y) := \max_{i=1, \dots, k} \rho(x_i, y_i)$ we have*

$$D^{(q)}(k \log N(E, \rho, \varepsilon) \mid X, \rho^k, s) \leq \varepsilon.$$

Essentially the same technique is applied in the proof of the next lemma. The result is comparable, but slightly more precise. This version is used to encode the jump positions.

Lemma 22 *Let Y be any random variable in $[0, 1]^k$. Then, for all $s > 0$, $r \geq 0$,*

$$D^{(q)}(r \mid Y, \|\cdot\|_\infty, s) \leq e^{-r/k}.$$

If Y is such that $Y_1 \leq \dots \leq Y_k$ almost surely then we can restrict ourselves to codebooks \mathcal{C} with $\hat{Y}_1 \leq \dots \leq \hat{Y}_k$ for all $\hat{Y} \in \mathcal{C}$.

Note that this may be a fairly weak estimate in concrete cases; however, it holds for all $k \geq 1$ and all $r \geq 0$. If more is known about the distribution of Y , much better (asymptotic) estimates are available, cf. [13], e.g. Theorem 6.2.

Proof: Let us first consider the case $e^r = n = ((m+1)/2)^k$ with $m \geq 1$. Then we can use a simple product quantizer. Namely, we set

$$\mathcal{C} := \{(y_1/m, \dots, y_k/m) \in [0, 1]^k : y_i \in \{1, 3, 5, \dots\}, i = 1, \dots, k\}.$$

Then $\#\mathcal{C} \leq ((m+1)/2)^k$ and thus

$$D^{(q)}(\log((m+1)/2)^k \mid Y, \|\cdot\|_\infty, s) \leq \left(\mathbb{E} \min_{a \in \mathcal{C}} \|Y - a\|_\infty^s \right)^{1/s} \leq m^{-1}.$$

For any $r > 0$ with $e^r \geq 2^k$, there exists an $m \geq 1$ such that $(\frac{m+1}{2})^k \leq e^r < (\frac{m+2}{2})^k$. Then

$$\begin{aligned} D^{(q)}(r \mid Y, \|\cdot\|_\infty, s) &\leq D^{(q)}(\log((m+1)/2)^k \mid Y, \|\cdot\|_\infty, s) \\ &\leq m^{-1} \leq (2e^{r/k} - 2)^{-1} \leq e^{-r/k}, \end{aligned}$$

where we used $e^r \geq 2^k$ in the last step.

Finally, for $1 \leq e^r \leq 2^k$,

$$D^{(q)}(r \mid Y, \|\cdot\|_\infty, s) \leq (\mathbb{E} \|Y - (1/2, \dots, 1/2)\|_\infty^s)^{1/s} \leq 1/2 \leq e^{-r/k}.$$

□

The last lemma can be strengthened if it is known that the random vector $Y = (Y_1, \dots, Y_k)$ satisfies $Y_1 \leq \dots \leq Y_k$.

Lemma 23 *There are absolute constants $c^*, \kappa > 0$ such that, for any random variable Y in $[0, 1]^k$ such that almost surely $Y_1 \leq \dots \leq Y_k$, we have, for all $s > 0$,*

$$D^{(q)}(r \mid Y, \|\cdot\|_\infty, s) \leq \frac{\kappa}{k} e^{-r/k}, \quad \text{for all } r \geq c^* k. \quad (12)$$

Proof: Let $m \geq k$ and consider

$$\mathcal{C} := \left\{ \left(\frac{y_1}{m}, \dots, \frac{y_k}{m} \right) : 1 \leq y_1 \leq y_2 \leq \dots \leq y_k \leq m, y_i \in \{1, \dots, m\} \right\}.$$

Clearly,

$$\#\mathcal{C} = \binom{m+k-1}{k}.$$

Note that for any $y \in [0, 1]^k$ with $y_1 \leq \dots \leq y_k$ we have $\min_{a \in \mathcal{C}} \|y - a\|_\infty \leq 1/m$. Thus,

$$D^{(q)}(\log(\#\mathcal{C}) \mid Y, \|\cdot\|_\infty, s) \leq \frac{1}{m},$$

for any random variable Y that satisfies the assumption of the lemma.

Note that, by Stirling's Formula, for some absolute constants $C_1, C_2, c^* > 0$,

$$\binom{k+1+k-1}{k} = \frac{(2k)!}{(k!)^2} \leq C_1 \frac{2^{2k} k^{2k} e^{-2k} \sqrt{2\pi 2k}}{k^{2k} e^{-2k} 2\pi k} \leq C_2 2^{2k} \leq e^{c^* k}. \quad (13)$$

Let $r \geq c^*k$. Then there is an $m \geq k$ such that

$$\binom{m+k-1}{k} < e^r \leq \binom{m+1+k-1}{k}, \quad (14)$$

because, as seen in (13),

$$\min_{m \geq k} \binom{m+k-1}{k} = \binom{k+k-1}{k} < \binom{k+1+k-1}{k} \leq e^{c^*k} \leq e^r.$$

Thus,

$$D^{(q)}(r \mid Y, \|\cdot\|_\infty, s) \leq D^{(q)}\left(\log \binom{m+k-1}{k} \mid Y, \|\cdot\|_\infty, s\right) \leq \frac{1}{m}. \quad (15)$$

By (14) and Stirling's Formula, for some absolute constants $C_3, C_4 > 0$,

$$\begin{aligned} e^r &\leq \binom{m+1+k-1}{k} = \frac{(m+k)!}{m!k!} \\ &\leq C_3 \frac{(m+k)^m (m+k)^k e^{-m-k} \sqrt{2\pi(m+k)}}{m^m e^{-m} \sqrt{2\pi m} k^k e^{-k} \sqrt{2\pi k}} \\ &\leq C_4 \left(1 + \frac{k}{m}\right)^m \frac{(2m)^k}{k^k} \sqrt{\frac{m+k}{mk}}. \end{aligned} \quad (16)$$

Observe that $\left(1 + \frac{k}{m}\right)^m \leq e^k$ and that $\frac{m+k}{mk} = \frac{1}{k} + \frac{1}{m} \leq 2$, for all m and k . Therefore, the term in (16) can be estimated by

$$C_5 (2e)^k \frac{m^k}{k^k} \leq \kappa^k \frac{m^k}{k^k},$$

where κ is an absolute constant. This implies $ke^{r/k} \leq \kappa m$ or $1/m \leq \kappa e^{-r/k}/k$. We deduce from (15) that for any $r \geq c^*k$ (12) holds, as asserted. \square

Now we can proceed with the proof of our first main result.

Proof of Theorem 1: Let X_k be a random variable that has the distribution of X conditioned upon the event that $N_X = k$, i.e. X that has k jumps. Let Y be the vector in $[0, 1]^k$ with the jump positions of X_k (in increasing order) and Z be the E^{k+1} -vector containing values of the process X_k between the jumps (in the order corresponding to when they occur), i.e. the initial value and the k jump destinations. Note that we can reconstruct X_k completely from the vectors Y and Z . Thus, it is sufficient to find good codebooks for Z and Y .

Let $\delta > 0$. By assumption, there is an $\varepsilon_0 = \varepsilon_0(\delta) \in]0, 1[$ such that for all $0 < \varepsilon \leq \varepsilon_0$,

$$\log N(E, \rho, \varepsilon) \leq (\gamma + \delta) \log 1/\varepsilon. \quad (17)$$

Let $r \geq \log 1/\varepsilon_0$. For $0 \leq k \leq k_0 := k_0(\delta, r) := r \min(1, (\log 1/\varepsilon_0(\delta))^{-1}) - 1$, let \mathcal{C}_k'' be a codebook for Z in (E^{k+1}, ρ^{k+1}) with

$$\left(\mathbb{E} \min_{\hat{Z} \in \mathcal{C}_k''} \rho^{k+1}(Z, \hat{Z})^s \right)^{1/s} \leq 2e^{-r/(k+1)}. \quad (18)$$

By Remark 21, \mathcal{C}_k'' can be chosen such that

$$\log \#\mathcal{C}_k'' \leq (k+1) \log N(E, \rho, e^{-r/(k+1)}) \leq (k+1)(\gamma + \delta) \log(e^{r/(k+1)}) = (\gamma + \delta)r, \quad (19)$$

where we used (17) and the choice of k_0 .

For $1 \leq k \leq k_0$, let \mathcal{C}_k' be a codebook for Y in $(\mathbb{R}^k, \|\cdot\|_\infty)$ with

$$\left(\mathbb{E} \min_{\hat{Y} \in \mathcal{C}_k'} \|Y - \hat{Y}\|_\infty^s \right)^{1/s} \leq 2(e^{r-k})^{-1/k}. \quad (20)$$

By Lemma 22, \mathcal{C}_k' can be chosen such that $\log \#\mathcal{C}_k' \leq r - k$.

Define $\mathcal{C}_0 := \mathcal{C}_0''$. For $k \neq 0$, let \mathcal{C}_k be the Cartesian product of the codebooks \mathcal{C}_k' and \mathcal{C}_k'' . Then $\log \#\mathcal{C}_k \leq r - k + (\gamma + \delta)r$ for all $0 \leq k \leq k_0$.

Let us define the following notation: for any $\hat{Y} \in \mathcal{C}_k'$, we set

$$F := \bigcup_{i=1}^k [\hat{Y}_i, Y_i[\cup [Y_i, \hat{Y}_i[\subseteq [0, 1[. \quad (21)$$

Note that on $[0, 1[\setminus F$, X can be reconstructed up to the error given in (18). Furthermore, note that the Lebesgue measure of F is less than $k \|Y - \hat{Y}\|_\infty$.

With the help of this information, we can estimate the error of approximating by \mathcal{C}_k when $k \neq 0$:

$$\begin{aligned} \mathbb{E} \min_{a \in \mathcal{C}_k} \rho_{\mathbf{D}}(X_k, a)^s &= \mathbb{E} \min_{a \in \mathcal{C}_k} \left(\int_0^1 \rho(X_k(t), a(t)) dt \right)^s \\ &\leq C_s \mathbb{E} \min_{a \in \mathcal{C}_k} \left(\left(\int_F \dots dt \right)^s + \left(\int_{[0,1[\setminus F} \dots dt \right)^s \right) \\ &\leq C_s \mathbb{E} \min_{\hat{Y} \in \mathcal{C}_k'} \min_{\hat{Z} \in \mathcal{C}_k''} \left((wk \|Y - \hat{Y}\|_\infty)^s + (\rho^{k+1}(Z, \hat{Z}))^s \right) \\ &= C_s \left((kw)^s \mathbb{E} \min_{\hat{Y} \in \mathcal{C}_k'} \|Y - \hat{Y}\|_\infty^s + \mathbb{E} \min_{\hat{Z} \in \mathcal{C}_k''} \rho^{k+1}(Z, \hat{Z})^s \right) \\ &\leq C_s \left((2kw)^s (e^{r-k})^{-s/k} + 2^s e^{-sr/(k+1)} \right) \\ &\leq Dk^s e^{-rs/(k+1)}, \end{aligned} \quad (22)$$

$$\leq Dk^s e^{-rs/(k+1)}, \quad (23)$$

having used (20) and (18) in the last but one step, where $D := C_s 2^s ((ew)^s + 1)$.

We define the codebook $\mathcal{C} := \bigcup_{0 \leq k \leq k_0} \mathcal{C}_k$. Then

$$\#\mathcal{C} \leq \sum_{0 \leq k \leq k_0} e^{r-k+(\gamma+\delta)r} \leq e^{r+(\gamma+\delta)r} \sum_{k=0}^{\infty} e^{-k} \leq e^{r+(\gamma+\delta)r+1}.$$

Thus,

$$\begin{aligned} D^{(q)}((1+\gamma+\delta)r+1 \mid X, \rho_{\mathbf{D}}, s)^s &\leq \mathbb{E} \min_{a \in \mathcal{C}} \rho_{\mathbf{D}}(X, a)^s \\ &\leq \sum_{0 \leq k \leq k_0} \mathbb{P}(N_X = k) \mathbb{E} \min_{a \in \mathcal{C}_k} \rho_{\mathbf{D}}(X_k, a)^s + \sum_{k > k_0} \mathbb{P}(N_X = k) \mathbb{E} \min_{a \in \mathcal{C}_0} \rho_{\mathbf{D}}(X_k, a)^s. \end{aligned} \quad (24)$$

Using (2), (23), and the trivial fact that $\rho_{\mathbf{D}}(X_k, a) \leq w$, the last expression is seen to be less than

$$\begin{aligned} &K e^{-\lambda} \left(2^s e^{-sr} + \sum_{1 \leq k \leq k_0} \frac{\lambda^k}{k!} D k^s e^{-rs/(k+1)} + \sum_{k > k_0} \frac{\lambda^k}{k!} w^s \right) \\ &= K e^{-\lambda} \left(2^s e^{-sr} + D \sum_{1 \leq k \leq k_0} \frac{\lambda^k}{k!} k^s e^{-rs/(k+1)} + w^s \sum_{k > k_0} \frac{\lambda^k}{k!} e^{-rs/(k+1)} e^{rs/(k+1)} \right) \\ &\leq K e^{-\lambda} \left(2^s e^{-sr} + D \sum_{1 \leq k \leq k_0} \frac{\lambda^k}{k!} e^{ks} e^{-rs/(k+1)} + \left(\frac{w}{\varepsilon_0}\right)^s \sum_{k > k_0} \frac{\lambda^k}{k!} e^{-rs/(k+1)} \right) \\ &\leq K e^{-\lambda} 2^s e^{-sr} + C_{K,s,w,\lambda,\varepsilon_0(\delta)} \sum_{k=0}^{\infty} \frac{(e^s \lambda)^k}{k!} e^{-e^s \lambda} e^{-rs/(k+1)}. \end{aligned} \quad (25)$$

Recall from Lemma 19 that the exponential order of the sum, when $r \rightarrow \infty$, is

$$-\sqrt{2rs \log(rs)} \sim -\sqrt{2rs \log r}$$

and that the constant in front of it does not depend on r . The first term in (25) also has no influence. Thus, for any $\delta > 0$,

$$\limsup_{r \rightarrow \infty} \frac{\log D^{(q)}((1+\gamma+\delta)r+1 \mid X, \rho_{\mathbf{D}}, s)}{\sqrt{r \log r}} \leq -\sqrt{\frac{2}{s}}.$$

Therefore

$$\limsup_{r \rightarrow \infty} \frac{\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s)}{\sqrt{r \log r}} \leq -\sqrt{\frac{2}{s(1+\gamma+\delta)}},$$

which holds for any $\delta > 0$. Letting δ tend to 0 gives the assertion. \square

Proof of Theorem 2: First we treat part (a). Again we condition upon the event that k jumps occur. Let X_k be a random variable that has the distribution of X conditioned

upon the event that $N_X = k$, i.e. that X has k jumps. Let, as above, Y be the vector in $[0, 1]^k$ with the jump positions of X_k and Z be the E^{k+1} -vector containing the values of the process X_k between the jumps. Recall that one can reconstruct X_k from Y and Z , so it suffices to find good codebooks for Y and Z .

Let $c^*, \kappa > 0$ be the absolute constants from Lemma 23. Let $r \geq \lambda c^*$. Fix $k \geq 1$. Let \mathcal{C}'_k be a codebook for Y in $(\mathbb{R}^k, \|\cdot\|_\infty)$ with

$$\left(\mathbb{E} \min_{\hat{Y} \in \mathcal{C}'_k} \|Y - \hat{Y}\|_\infty^s \right)^{1/s} \leq \frac{2\kappa}{k} \left(e^{rk/\lambda} \right)^{-1/k} = \frac{2\kappa}{k} e^{-r/\lambda}.$$

By Lemma 23 and the fact that $Y_1 \leq \dots \leq Y_k$, \mathcal{C}'_k can be chosen such that $\log \#\mathcal{C}'_k \leq kr/\lambda$.

Furthermore, for $k \geq 0$, let \mathcal{C}''_k be a codebook for Z in (E^{k+1}, ρ^{k+1}) with

$$\left(\mathbb{E} \min_{\hat{Z} \in \mathcal{C}''_k} \rho^{k+1}(Z, \hat{Z})^s \right)^{1/s} \leq 2e^{-r/\lambda}. \quad (26)$$

By Remark 21, \mathcal{C}''_k can be chosen such that

$$\log \#\mathcal{C}''_k \leq (k+1) \log N(E, \rho, e^{-r/\lambda}). \quad (27)$$

Let \mathcal{C}_k be the Cartesian product of the codebooks \mathcal{C}'_k and \mathcal{C}''_k . Then $\log \#\mathcal{C}_k \leq kr/\lambda + (k+1) \log N(E, \rho, e^{-r/\lambda})$.

Let F be defined as in (21). In case k jumps occur ($k \neq 0$) we approximate X by a function from \mathcal{C}_k , which gives an error of at most

$$\begin{aligned} \mathbb{E} \min_{a \in \mathcal{C}_k} \rho_{\mathbf{D}}(X_k, a)^s &= \mathbb{E} \min_{a \in \mathcal{C}_k} \left(\int_0^1 \rho(X_k(t), a(t)) dt \right)^s \\ &\leq C_s \mathbb{E} \min_{a \in \mathcal{C}_k} \left(\left(\int_F \dots dt \right)^s + \left(\int_{[0,1] \setminus F} \dots dt \right)^s \right) \\ &\leq C_s \mathbb{E} \min_{\hat{Y} \in \mathcal{C}'_k} \min_{\hat{Z} \in \mathcal{C}''_k} \left((wk \|Y - \hat{Y}\|_\infty)^s + (\rho^{k+1}(Z, \hat{Z}))^s \right) \end{aligned} \quad (28)$$

$$\begin{aligned} &= C_s \left((kw)^s \mathbb{E} \min_{\hat{Y} \in \mathcal{C}'_k} \|Y - \hat{Y}\|_\infty^s + \mathbb{E} \min_{\hat{Z} \in \mathcal{C}''_k} \rho^{k+1}(Z, \hat{Z})^s \right) \\ &\leq C_s 2^s \left((w\kappa)^s e^{-rs/\lambda} + e^{-sr/\lambda} \right) \\ &\leq C_s 2^s ((w\kappa)^s + 1) e^{-rs/\lambda}. \end{aligned} \quad (29)$$

For $k = 0$, set $\mathcal{C}_0 := \mathcal{C}''_0$. Then the error is less than $2e^{-rs/\lambda}$, by (26).

On the other hand, this procedure has an expected nat length of at most

$$\begin{aligned} K e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \log(\#\mathcal{C}_k) &= K \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \left(kr/\lambda + (k+1) \log N(E, \rho, e^{-r/\lambda}) \right) \\ &= K \left(r + (\lambda + 1) \log N(E, \rho, e^{-r/\lambda}) \right). \end{aligned}$$

Therefore, similarly to (24),

$$\begin{aligned} D^{(e)}(K(r + (\lambda + 1) \log N(E, \rho, e^{-r/\lambda})) \mid X, \rho_{\mathbf{D}}, s)^s \\ \leq K e^{-\lambda} \left(2e^{-rs/\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} C_s 2^s ((w\kappa)^s + 1) e^{-rs/\lambda} \right) \\ = K \left(2e^{-rs/\lambda} e^{-\lambda} + C_s 2^s ((w\kappa)^s + 1) e^{-rs/\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \right) \leq K C'_s (w^s + 1) e^{-rs/\lambda}. \quad (30) \end{aligned}$$

where C'_s only depends on s . This yields the assertion (a).

To see (b) one only has to recall that in case the jump positions are distributed as a Poisson point process we can choose $K = 1$ in (2).

Let us finally show (c). In the case of a discrete space $E = \{x_1, \dots, x_q\}$ with $w = \max_{x,y \in E} \rho(x, y)$, we can choose $\log \#\mathcal{C}_k'' = (k+1) \log q$. Thus, on $[0, 1] \setminus F$, no error arises. This allows to replace the right-hand side in (29) by $(2w\kappa)^s e^{-rs/\lambda}$. Therefore, the upper bound in (30) becomes $K(2w\kappa)^s (1 - e^{-\lambda}) e^{-rs/\lambda}$, where $\kappa > 0$ is the absolute constant from Lemma 23. This finishes the proof of (c). \square

Note that no assumption is necessary on the correlation of the jump positions and increments.

Let us now indicate the changes that are necessary to prove Theorem 3.

Proof of Theorem 3: The proof carries over almost literally from Theorems 1 and 2, respectively. The only differences concern the assumption on $d^{(q)}$ instead of the metric entropy N , the fixed initial position, and the possibly unbounded jumps.

In this case, we encode the increments instead of the jump destinations. Let Y be as above, but Z denote the E^k vector with the increments, i.e. $Z := (Z^{(1)}, \dots, Z^{(k)})$ with $Z^{(i)} := X(Y_i) - X(Y_i^-)$. Note that we can reconstruct X from Y and Z , since we assumed $X(0)$ to be deterministic.

The first change is to replace (19) by

$$\log \#\mathcal{C}_k'' \leq k \log d^{(q)}(e^{-r/k} \mid Z^{(1)}, \rho, s) \leq k(\gamma + \delta) \log e^{r/k} = (\gamma + \delta)r$$

in the proof for the quantization error. For the entropy coding error one has to replace

(27) by

$$\log \#\mathcal{C}_k'' \leq k \log d^{(a)}(e^{-r/\lambda} \mid Z^{(1)}, \rho, s).$$

The second issue concerns a certain refinement in order to deal with the possibly unbounded jumps. Here, we need that we deal with a normed space. We will show that, on average, the high jumps do not have any influence on the rate. In fact, the only modification affects (22), where we estimate by

$$C_s \mathbb{E} \min_{\hat{Y} \in \mathcal{C}'_k} \min_{\hat{Z} \in \mathcal{C}''_k} \left[\left(\max_{1 \leq n, m \leq k} \left\| \sum_{i=1}^n Z^{(i)} - \sum_{i=1}^m \hat{Z}^{(i)} \right\| k \|Y - \hat{Y}\|_\infty \right)^s + \left(k \rho^k(Z, \hat{Z}) \right)^s \right],$$

which is required due to the fact that we cannot estimate by a finite diameter w (modification in the first term) and the errors may add up over all the jumps, since we encode the increments and not the absolute positions (modification in the second term).

The first term can be estimated by

$$\begin{aligned} C_s \mathbb{E} \min_{\hat{Y} \in \mathcal{C}'_k} \min_{\hat{Z} \in \mathcal{C}''_k} & \left(\sum_{i=1}^k \|Z^{(i)} - \hat{Z}^{(i)}\| + \sum_{i=1}^k \|Z^{(i)}\| \right)^s k^s \|Y - \hat{Y}\|_\infty^s \\ & \leq C_s^2 \left(k^s \mathbb{E} \min_{\hat{Z} \in \mathcal{C}''_k} \rho^k(Z, \hat{Z})^s + k^{s+1} \mathbb{E} \|Z^{(1)}\|^s \right) k^s \mathbb{E} \min_{\hat{Y} \in \mathcal{C}'_k} \|Y - \hat{Y}\|_\infty^s \\ & \leq C_s^2 (2^s + \mathbb{E} \|Z^{(1)}\|^s) k^{2s+1} \mathbb{E} \min_{\hat{Y} \in \mathcal{C}'_k} \|Y - \hat{Y}\|_\infty^s, \end{aligned}$$

where the last step comes from (18).

This leads to an additional factor Ck^{s+1} in (23) which has no influence on the order. Note furthermore that this argument needs that the jump positions and the increments are independent (in order to separate the expectations) and that the increments are identically distributed (as $Z^{(1)}$). It is not needed that the increments are independent among each other.

Analogously, for the proof of the entropy coding error, (28) is modified, which leads to an additional factor of Ck^{s+1} in (29), which leaves the resulting order unchanged, but which *does* change the constant. \square

3 Lower bound for the quantization error

In this section, we prove the lower bounds for the quantization error. Essentially we employ a small ball argument, i.e. we construct an event of not too small probability that still leaves sufficient uncertainty for the error to be large.

First we prove Theorem 4.

Proof of Theorem 4: Let us fix $k > 0$ and $\delta > 0$ (to be chosen later) and define intervals $I_j := \left[\frac{j-1}{k} + \frac{1}{4k}, \frac{j}{k} - \frac{1}{4k} \right]$, $j = 1, \dots, k$. Note that $\lambda_1(I_j) = 1/(2k)$. Let A be the event that X has exactly k jumps at Y_1, \dots, Y_k , such that $Y_j \in I_j$, for all $j = 1, \dots, k$, and that the moduli of the increments are all greater than ε_0 . Since the Y_i and Z_i are independent (by condition (*)) and the Y_i are distributed according to a Poisson point process, we have

$$\begin{aligned} \mathbb{P}(A) &= \prod_{j=1}^k \mathbb{P}(\text{exactly one jump in } I_j, Z_j > \varepsilon_0) \cdot \mathbb{P}(\text{no jump in } [(j-1)/k, j/k] \setminus I_j) \\ &\geq \prod_{j=1}^k \left(\frac{\lambda}{2k} e^{-\lambda/(2k)} \delta_0 \cdot e^{-\lambda/(2k)} \right) = \left(\frac{\delta_0 \lambda}{2k} \right)^k e^{-\lambda}. \end{aligned}$$

Step 1: Let X_A be a random variable with the distribution of X conditioned upon the event A . Then

$$\begin{aligned} D^{(q)}(r \mid X_A, \rho_{\mathbf{D}}, s)^s &= \inf_{\log(\#\mathcal{C}) \leq r} \mathbb{E}_{X_A} \min_{f \in \mathcal{C}} \rho_{\mathbf{D}}(X_A, f)^s \\ &\geq \inf_{\log(\#\mathcal{C}) \leq r} \delta^s \mathbb{P}(\forall f \in \mathcal{C} : \rho_{\mathbf{D}}(X_A, f) \geq \delta) \\ &= \inf_{\log(\#\mathcal{C}) \leq r} \delta^s (1 - \mathbb{P}(\exists f \in \mathcal{C} : \rho_{\mathbf{D}}(X_A, f) < \delta)) \\ &\geq \inf_{\log(\#\mathcal{C}) \leq r} \delta^s \left(1 - (\#\mathcal{C}) \sup_f \mathbb{P}(\rho_{\mathbf{D}}(X_A, f) < \delta) \right) \\ &\geq \delta^s \left(1 - e^r \sup_f \mathbb{P}(\rho_{\mathbf{D}}(X_A, f) < \delta) \right), \end{aligned} \tag{31}$$

where the supremum is taken over all functions f in $\mathbf{D}([0, 1[, E)$. For such f , we have

$$\begin{aligned} \mathbb{P}(\rho_{\mathbf{D}}(X_A, f) < \delta) &= \mathbb{P}\left(\int_0^1 \rho(X_A(t), f(t)) dt < \delta\right) \\ &\leq \mathbb{P}\left(\bigcap_{j=1}^k \left\{ \int_{I_j} \rho(X_A(t), f(t)) dt < \delta \right\}\right) = \mathbb{E} \mathbb{P}\left(\bigcap_{j=1}^k \left\{ \int_{I_j} \rho(X_A(t), f(t)) dt < \delta \right\} \middle| Z\right), \end{aligned} \tag{32}$$

where $Z = (X_A(0), X_A(Y_1), \dots, X_A(Y_k))$ is the vector with the jump destinations. By condition (*), we have that, conditioned upon Z , the events

$$\left(\left\{ \int_{I_j} \rho(X_A(t), f(t)) dt < \delta \right\} \right)_{j=1}^k$$

are independent, since each of them only depends on the jump position in the respective

interval. This together with (32) shows

$$\sup_f \mathbb{P}(\rho_{\mathbf{D}}(X_A, f) < \delta) \leq \sup_f \mathbb{E} \prod_{j=1}^k \mathbb{P}\left(\int_{I_j} \rho(X_A(t), f(t)) dt < \delta \middle| Z\right). \quad (33)$$

Step 2: Now we estimate each term in the product separately. Fix $j \in \{1, \dots, k\}$. Define $l_j := \frac{j-1}{k} + \frac{1}{4k}$, i.e. the left end point of the interval I_j . Furthermore, we define

$$B_j := \{t \in I_j : \rho(X_A(l_j), f(t)) < \varepsilon_0/2\} \quad \text{and} \quad C_j := \{t \in I_j : X_A(t) = X_A(l_j)\}.$$

Then we show that

$$\int_{I_j} \rho(X_A(t), f(t)) dt < \delta \quad \Rightarrow \quad \lambda_1(B_j \Delta C_j) < \frac{2\delta}{\varepsilon_0}, \quad (34)$$

where $B_j \Delta C_j := (B_j^c \cap C_j) \cup (B_j \cap C_j^c)$. Indeed, assume that we had $\lambda_1(B_j \Delta C_j) \geq 2\delta/\varepsilon_0$. Then

$$\begin{aligned} & \int_{I_j} \rho(X_A(t), f(t)) dt \\ & \geq \int_{B_j^c \cap C_j} \rho(X_A(t), f(t)) dt + \int_{B_j \cap C_j^c} \rho(X_A(t), f(t)) dt \geq \frac{\varepsilon_0}{2} \lambda_1(B_j \Delta C_j) \geq \delta, \end{aligned}$$

where we used the triangle inequality in the last but one step. This shows (34); and we thus have

$$\mathbb{P}\left(\int_{I_j} \rho(X_A(t), f(t)) dt < \delta \middle| Z\right) \leq \mathbb{P}\left(|\lambda_1(B_j) - \lambda_1(C_j)| < \frac{2\delta}{\varepsilon_0} \middle| Z\right).$$

Note that, conditioned upon Z , $\lambda_1(B_j)$ is a deterministic value (depending on $X_A(l_j)$ and f), whereas $\lambda_1(C_j)$ is a random variable that is uniformly distributed in $[0, 1/(2k)]$, since the point in I_j where the jump of X_A occurs is uniformly distributed in I_j . Therefore,

$$\mathbb{P}\left(|\lambda_1(B_j) - \lambda_1(C_j)| < \frac{2\delta}{\varepsilon_0} \middle| Z\right) \leq \frac{8\delta k}{\varepsilon_0}.$$

Step 3: This shows, continuing (33), that $\sup_f \mathbb{P}(\rho_{\mathbf{D}}(X_A, f) < \delta) \leq (8k\delta/\varepsilon_0)^k$. Substituting this estimate back into (31), we obtain

$$D^{(q)}(r \mid X_A, \rho_{\mathbf{D}}, s)^s \geq \delta^s \left(1 - e^r (8\delta k/\varepsilon_0)^k\right).$$

Therefore,

$$D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s)^s \geq \mathbb{P}(A) \cdot D^{(q)}(r \mid X_A, \rho_{\mathbf{D}}, s)^s \geq \left(\frac{\delta_0 \lambda}{2k}\right)^k e^{-\lambda \delta^s} \left(1 - e^r \left(\frac{8k\delta}{\varepsilon_0}\right)^k\right).$$

Now we can optimize $k \geq 1$ and $\delta > 0$ to obtain the largest possible lower bound. We set

$$\delta := \frac{\varepsilon_0}{8k} \left(\frac{1}{2} e^{-r} \right)^{1/k}.$$

Then the last estimate becomes

$$D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \geq \left(\frac{\delta_0 \lambda}{2k} \right)^{k/s} e^{-\lambda/s \delta} 2^{-1/s}.$$

We set

$$k := \lfloor \sqrt{2sr / \log r} \rfloor \sim \sqrt{2sr / \log r}.$$

Taking logarithms of the last estimate shows that

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \lesssim \frac{k}{s} \log k + r/k \sim \sqrt{\frac{2}{s}} r \log r,$$

as asserted. □

The proof of Theorem 14 contains the same idea as the one of Theorem 4 and carries over almost literally. Therefore, we only indicate the necessary changes.

Proof of Theorem 14: By assumption, $Z^{(1)}$ has an absolutely continuous component. Let $S \subseteq \mathbb{R}^d$ be a measurable set with $\lambda_d(S) > 0$ on which $Z^{(1)}$ has a positive bounded density w.r.t. the Lebesgue measure and such that $0 \notin S$. Define $\varepsilon_0 := \text{dist}(S, 0)/2 > 0$.

This time, A is defined as follows: let A be the event that X has exactly k jumps at Y_1, \dots, Y_k , such that $Y_j \in I_j$, for all $j = 1, \dots, k$, and that the corresponding increments (i.e. $Z^{(j)} = X(Y_j) - X(Y_{j-})$) are of a height in S . Due to the Poissonian nature of the point process and since increments and positions are independent, we have

$$\begin{aligned} \mathbb{P}(A) &= \prod_{j=1}^k \mathbb{P}(\text{exactly one jump in } I_j) \cdot \mathbb{P}(X(Y_j) - X(Y_{j-}) \in S) \cdot \\ &\quad \cdot \mathbb{P}(\text{no jump in } [(j-1)/k, j/k] \setminus I_j) \\ &= \prod_{j=1}^k \left(\frac{\lambda}{2k} e^{-\lambda/(2k)} \cdot q_S \cdot e^{-\lambda/(2k)} \right) = \left(\frac{\lambda q_S}{2k} \right)^k e^{-\lambda}, \end{aligned}$$

where $q_S := \mathbb{P}(Z^{(1)} \in S) > 0$. Regarding (31), the proof is analogous to that of Theorem 4. We set $Z = (X_A(Y_1), \dots, X_A(Y_k))$ for the vector with the jump destinations. In (32)

and (33) we estimate a bit more carefully and obtain:

$$\begin{aligned} & \mathbb{P}(\rho_{\mathbf{D}}(X_A, f) < \delta) \\ & \leq \mathbb{E} \prod_{j=1}^k \mathbb{P} \left(\int_{I_j} \rho(X_A(t), f(t)) dt < \delta, \int_{(4j-1)/(4k)}^{j/k} \rho(X_A(t), f(t)) dt < \delta \middle| Z \right). \end{aligned}$$

As in the proof of Theorem 4, the sets B_j and C_j are introduced and (34) is established. Let $r'_j := j/k$ and $r_j := r'_j - 1/(4k)$. Because of (34) and since $X_A(t) = X_A(r'_j) = X_A(r_j)$ on $[r_j, r'_j]$, the last expression is less than

$$\mathbb{E} \prod_{j=1}^k \mathbb{P} \left(\lambda_d(B_j \Delta C_j) < \delta, \int_{r_j}^{r'_j} \rho(X_A(r_j), f(t)) dt < \delta \middle| Z \right).$$

Note that, conditioned upon Z , the events $\lambda_d(B_j \Delta C_j) < \delta$ and $\int_{r_j}^{r'_j} \rho(X_A(r_j), f(t)) dt < \delta$ are independent, since the second event only depends on Z , i.e. it is deterministic. Thus the last expression equals

$$\mathbb{E} \prod_{j=1}^k \mathbb{P}(\lambda_d(B \Delta C) < \delta | Z) \mathbb{P} \left(\int_{r_j}^{r'_j} \rho(X_A(r_j), f(t)) dt < \delta \middle| Z \right).$$

The first term can be estimated as in the proof of Theorem 4 by $8\delta k/\varepsilon_0$, which allows to estimate the last expression by

$$\left(\frac{8\delta k}{\varepsilon_0} \right)^k \mathbb{E} \prod_{j=1}^k \mathbb{P} \left(\int_{r_j}^{r'_j} \rho(X_A(r_j), f(t)) dt < \delta \middle| Z \right).$$

In order to treat the second term, note that it equals

$$\begin{aligned} & \mathbb{P} \left(\int_{r_j}^{r'_j} \rho(X_A(r_j), f(t)) dt < \delta, j = 1, \dots, k \right) \\ & = \mathbb{E} \mathbb{P} \left(\int_{r_j}^{r'_j} \rho(X_A(r_j), f(t)) dt < \delta, j = 1, \dots, k \middle| Z^{(1)}, \dots, Z^{(k-1)} \right). \quad (35) \end{aligned}$$

Note that the last condition (for $j = k$) is the only non-deterministic condition in the probability. It depends on $Z^{(k)}$, which is an \mathbb{R}^d -valued random variable distributed as $Z^{(1)}$.

By the definition of the event A , $Z^{(k)}$ attains values in S . Thus,

$$\begin{aligned} & \mathbb{P} \left(\int_{r_k}^{r'_k} \rho(X_A(r_k), f(t)) dt < \delta \mid Z^{(1)}, \dots, Z^{(k-1)} \right) \\ &= \mathbb{P} \left(\int_{r_k}^{r'_k} \left\| \sum_{j=1}^k Z^{(j)} - f(t) \right\|_{\infty} dt < \delta \mid Z^{(1)}, \dots, Z^{(k-1)} \right) \\ &\leq \mathbb{P} \left(\left\| \frac{Z^{(k)}}{4k} + \int_{r_k}^{r'_k} \sum_{j=1}^{k-1} Z^{(j)} - f(t) dt \right\|_{\infty} < \delta \mid Z^{(1)}, \dots, Z^{(k-1)} \right), \end{aligned}$$

where the integral is to be understood componentwise. Note that $\int_{r_k}^{r'_k} \sum_{j=1}^{k-1} Z^{(j)} - f(t) dt$ is a deterministic value in \mathbb{R}^d , conditioned upon $(Z^{(1)}, \dots, Z^{(k-1)})$. Thus, the last term is bounded from above by $\varepsilon'_0(8k\delta)^d$, where ε'_0 is the supremum of the density of $Z^{(k)} \stackrel{d}{=} Z^{(1)}$ in S . In the same way, successively the other terms can be reduced; and the expression in (35) can be estimated by $\varepsilon_0''(8k\delta)^{dk}$. Therefore,

$$\sup_f \mathbb{P}(\rho_{\mathbf{D}}(X_A, f) < \delta) \leq (k\delta\varepsilon_0'')^{k(1+d)},$$

where $\varepsilon_0'' = 8 \min(1/\varepsilon_0, \varepsilon'_0)$. Continuing as in Step 3 of the proof of Theorem 4 shows

$$D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s)^s \geq \mathbb{P}(A) \cdot D^{(q)}(r \mid X_A, \rho_{\mathbf{D}}, s)^s \geq \left(\frac{\lambda q_S}{2k} \right)^k e^{-\lambda\delta^s} (1 - e^r (k\delta\varepsilon_0'')^{k(1+d)}).$$

This time we set

$$\delta := \frac{1}{\varepsilon_0'' k} \left(\frac{1}{2} e^{-r} \right)^{1/(k(1+d))}.$$

Then again

$$D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \geq \left(\frac{\lambda q_S}{2k} \right)^{k/s} e^{-\lambda/s\delta} 2^{-1/s},$$

where this time we set

$$k := \left\lceil \sqrt{\frac{2s}{1+d} \frac{r}{\log r}} \right\rceil \sim \sqrt{\frac{2s}{1+d} \frac{r}{\log r}}.$$

This eventually leads to

$$-\log D^{(q)}(r \mid X, \rho_{\mathbf{D}}, s) \lesssim \frac{k}{s} \log k + r/(k(1+d)) \sim \sqrt{\frac{2}{s(1+d)}} r \log r,$$

as asserted. \square

4 Lower bound for the entropy coding error

In this section we prove a corresponding lower bound for the entropy coding error (in fact, for the distortion rate function) for a jump process where the underlying point process is Poissonian. We use the notation from [14], in particular, for the distortion rate function

$$D(r | X, \rho_D, s) := \inf \left\{ \left(\mathbb{E} \rho_{\mathbf{D}}(X, \hat{X})^s \right)^{1/s} : I(X; \hat{X}) \leq r \right\},$$

and the notion of mutual information:

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

We recall that $D(r | X, \rho, s) \leq D^{(e)}(r | X, \rho, s) \leq D^{(a)}(r | X, \rho, s)$ for any random variable, all moments and any distortion measure. Therefore, a lower bound for D immediately translates into a lower bound for $D^{(e)}$.

Let us state the assumptions of the main result of this section. We shall require that X is a jump process (on the index set $[0, 1[$) whose jumps form a Poisson process of intensity $\lambda > 0$. Furthermore, we assume that ρ defines a metric on E and that the moduli of the jumps of X are a.s. bounded from below by a constant $\varepsilon_0 > 0$.

As before, we denote by (Y_i) the jump times of the process X and by N_X the random number of jumps of X ; we set $Y_0 = 0$. Moreover, we assume that conditioned upon $N_X = k$ the random vector $(A_i) := (X(Y_i))_{i=0}^k$ and the jump times (Y_i) are independent (condition (*)). In the rest of this section, we prove the following stronger version of Theorem 7.

Theorem 24 *Under the above assumptions one has*

$$D(r | X, \rho_{\mathbf{D}}, 1) \geq \varepsilon_0 C \min(1, \lambda) e^{-r/\lambda},$$

where $C > 0$ is an absolute constant.

Let us shortly describe the idea of the proof. We relate the coding complexity of the jump process to that of the random jump times. Controlling the complexity of the jump times by using Shannon's lower bound then leads to a lower bound in terms of a variational problem. The proof is based on several lemmas and a particular random partition of $[0, 1[$.

We denote by \mathcal{D}_m the dyadic subintervals of $[0, 1[$ of the m -th level, that is

$$\mathcal{D}_m = \{(j-1)2^{-m}, j2^{-m}\} : j = 1, \dots, 2^m\}$$

We construct for any collection t_1, \dots, t_k of distinct points in $[0, 1[$ a finite binary tree as

follows. Let $\mu = \sum_{i=1}^k \delta_{t_k}$ with δ_z denoting the Dirac mass in z . The root of the tree will be associated with the interval $[0, 1[$ and it will be marked by the number $\mu([0, 1[) = k$. If $k \in \{0, 1\}$, then the construction ends and the root is also a leaf of the tree. If $k \geq 2$, the root of the tree is attached two children namely the two dyadic intervals of \mathcal{D}_1 that are contained in $[0, 1[$: $[0, 1/2[$ and $[1/2, 1[$. Again we mark each of the nodes with their corresponding masses. Each node that has mass 0 or 1 becomes a leaf of the tree, and for each node with mass greater than 1 we attach the two dyadic intervals of the next level that are contained in the interval and we continue in analogy to above.

By the construction, each leaf contains either one or no point. We shall denote by $\pi_k(t_1, \dots, t_k) := (I_1, \dots, I_k)$ the k -intervals associated to the leaves with positive mass. In order to make the definition unique we arrange the intervals in their natural order.

Lemma 25 *Let $k \geq 1$ and $(I_1, \dots, I_k) \in \text{im}(\pi_k)$. Conditioned upon the event $\{N_X = k, \pi_k(Y_1, \dots, Y_k) = (I_1, \dots, I_k)\}$ we have that $(Y_1, \dots, Y_k) \stackrel{d}{=} (U_1, \dots, U_k)$, where U_i are independent random variables that are uniformly distributed on I_i , respectively.*

Proof: First note that for any collection of distinct points $t_1, \dots, t_k \in [0, 1[$ such that $\sum_{j=1}^k \mathbb{1}_{I_i}(t_j) = 1$ for all $i = 1, \dots, k$ one retrieves $\pi_k(t_1, \dots, t_k) = (I_1, \dots, I_k)$. On the other hand, any collection of points which yields $\sum_{j=1}^k \mathbb{1}_{I_i}(t_j) \neq 1$ for one i , induces a different tree and $\pi_k(t_1, \dots, t_k) \neq (I_1, \dots, I_k)$.

Therefore, the following two events coincide

$$\{N_X = k, \pi_k(Y_1, \dots, Y_k) = (I_1, \dots, I_k)\} = \left\{N_X = k, \sum_{j=1}^k \mathbb{1}_{I_i}(Y_j) = 1 \text{ for } i = 1, \dots, k\right\}.$$

Recall that the times (Y_i) form a Poisson process on $[0, 1[$ so that conditioned on $\{N_X = k, \pi(Y_1, \dots, Y_k) = (I_1, \dots, I_k)\}$ one has $(Y_1, \dots, Y_k) \stackrel{d}{=} (U_1, \dots, U_k)$, where U_i are independent random variables uniformly distributed on I_i . \square

Lemma 26 *Fix $k \geq 1$, $(I_1, \dots, I_k) \in \text{im}(\pi_k)$ and distinct points $a_0, \dots, a_k \in E$ with $|a_i - a_{i-1}| \geq \varepsilon_0$ for $i = 1, \dots, k$. Moreover, let μ denote the distribution of a process in $\mathbf{D}([0, 1[, E)$ that has jump positions at k uniformly distributed times in the intervals I_1, \dots, I_k and that attains the values a_0, \dots, a_k in the given order. Then*

$$D(r \mid \mu, \rho_{\mathbf{D}}, 1) \geq \varepsilon_0 \frac{k}{2e} \left(\prod_{i=1}^k |I_i| \right)^{1/k} e^{-r/k}.$$

Proof: With slight abuse of notation we shall denote by $X = (X(t))_{t \in [0, 1[}$ a μ -distributed process and we let Y_1, \dots, Y_k denote the ordered k jump positions of X . Due to Lemma 25 the times Y_1, \dots, Y_k are independent and each Y_i is uniformly distributed on I_i .

Now let $\hat{X} = (\hat{X}(t))_{t \in [0,1[}$ denote a $\mathbf{D}([0,1[, E)$ -valued reconstruction with $I(X; \hat{X}) \leq r$. We define $X_t^i = a_{i-1}$ for $t < Y_i$ and $X_t^i = a_i$ for $t \geq Y_i$. Also we set $\hat{X}_t^i = \hat{X}(t)$ for $t \in I_i$ and $\hat{X}_t^i = X_t^i$ otherwise. Then clearly

$$\rho_{\mathbf{D}}(X, \hat{X}) \geq \sum_{i=1}^k \rho_{\mathbf{D}}(X^i, \hat{X}^i).$$

Next, we will provide a lower bound for the right hand side in the latter inequality.

For each fixed $i = 1, \dots, k$ we define ν_i to be the probability kernel of the regular conditional probability $\mathbb{P}(Y_i \in \cdot | \hat{X} = \cdot)$. Next we choose $\hat{Y}_i = \hat{Y}_i(\hat{X})$ to be the first time $t \in [0, 1[$ for which the probability $\nu_i(\hat{X}, [0, t])$ is greater or equal to the threshold $1/2$.

We observe that for $t \in [0, 1[$

$$\begin{aligned} \mathbb{E}[\rho(X_t^i, \hat{X}_t^i) | \hat{X}] &\geq \mathbb{P}(X_t^i = a_{i-1} | \hat{X}) \wedge \mathbb{P}(X_t^i = a_i | \hat{X}) \left[\rho(a_{i-1}, \hat{X}_t^i) + \rho(a_i, \hat{X}_t^i) \right] \\ &\geq \mathbb{P}(X_t^i = a_{i-1} | \hat{X}) \wedge \mathbb{P}(X_t^i = a_i | \hat{X}) \rho(a_{i-1}, a_i) \\ &= \varepsilon_0 \mathbb{P}(Y_i < t | \hat{X}) \wedge \mathbb{P}(Y_i \geq t | \hat{X}). \end{aligned}$$

Consequently, the approximation error satisfies

$$\begin{aligned} \mathbb{E}[\rho_{\mathbf{D}}(X^i, \hat{X}^i) | \hat{X}] &\geq \varepsilon_0 \int_0^1 \mathbb{P}(Y_i < t | \hat{X}) \wedge \mathbb{P}(Y_i \geq t | \hat{X}) dt \\ &= \varepsilon_0 \left[\int_0^{\hat{Y}_i} \mathbb{E}[\mathbb{1}_{\{Y_i < t\}} | \hat{X}] dt + \int_{\hat{Y}_i}^1 \mathbb{E}[\mathbb{1}_{\{Y_i \geq t\}} | \hat{X}] dt \right] \\ &= \varepsilon_0 \mathbb{E}(|Y_i - \hat{Y}_i| | \hat{X}) \end{aligned}$$

and one gets

$$\mathbb{E}[\rho_{\mathbf{D}}(X, \hat{X})] \geq \varepsilon_0 \sum_{i=1}^k \mathbb{E}|Y_i - \hat{Y}_i|. \quad (36)$$

We shall now use the Shannon lower bound to derive a lower bound for the right hand side of the latter equation. For ease of notation we write shortly $Y = (Y_1, \dots, Y_k)$ and $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_k)$. We need the notation for the continuous entropy and its conditional counterpart: for \mathbb{R}^k -valued random vectors Z and \hat{Z} we denote

$$h(Z) := - \int \log \frac{d\mathbb{P}_Z}{d\lambda_k} d\mathbb{P}_Z \quad \text{and} \quad h(Z | \hat{Z}) := - \int \log \frac{d\mathbb{P}_{Z|\hat{Z}}}{d\lambda_k} d\mathbb{P}_{Z, \hat{Z}},$$

provided the Radon-Nikodym derivatives exist and the integrals are well-defined.

Since \hat{Y} is $\sigma(\hat{X})$ -measurable we have $I(Y; \hat{Y}) \leq I(X; \hat{X}) \leq r$; so that by the Shannon lower bound

$$\begin{aligned} r &\geq I(Y; \hat{Y}) = h(Y) - h(Y|\hat{Y}) \\ &= h(Y) - h(Y - \hat{Y}|\hat{Y}) \geq h(Y) - h(Y - \hat{Y}). \end{aligned}$$

In particular, $Y - \hat{Y}$ is absolutely continuous and its differential entropy is well-defined. Next, we set $d := \mathbb{E}\|Y - \hat{Y}\|_{\ell_1^k}$ and estimate the term $h(Y - \hat{Y})$ from above by

$$\phi(d) = \sup_{\substack{Z, \mathbb{P}_Z \ll \lambda_k \\ \mathbb{E}\|Z\|_{\ell_1^k} \leq d}} h(Z).$$

Using Lemma 6.4 from [1] (which is based on ideas from [8]) one can easily show that

$$\sup_{Z, \mathbb{P}_Z \ll \lambda_k, \sum_{i=1}^k \mathbb{E}|Z_i| \leq d} h(Z) = \sum_{i=1}^k \log \left(\frac{2de}{k} \right).$$

Consequently, $r \geq h(Y) - k \log(2ed/k)$ or, equivalently,

$$d \geq \frac{k}{2e} e^{h(Y)/k} e^{-r/k}.$$

Moreover, the entropy of Y satisfies

$$h(Y) = \sum_{i=1}^k h(Y_i) = \sum_{i=1}^k \log \frac{1}{|I_i|}.$$

and we conclude that

$$d \geq \frac{k}{2e} \left(\prod_{i=1}^k |I_i| \right)^{1/k} e^{-r/k},$$

which together with (36) shows the assertion. \square

A crucial quantity in the latter lower bound for the distortion rate function is the length of the intervals I_i . Later we will use the following estimate:

Lemma 27 *Let $t_1, \dots, t_k \in [0, 1[$ denote k distinct points ordered by their size and let $(I_1, \dots, I_k) = \pi_k(t_1, \dots, t_k)$. With $t_0 = -\infty$ and $t_{k+1} = \infty$ we get for each $i = 1, \dots, k$ that*

$$|I_i| \geq \frac{1}{2} (t_i - t_{i-1}) \wedge (t_{i+1} - t_i).$$

Proof: By definition I_i is the largest dyadic interval that only contains the point t_i and the assertion follows since all half-open intervals of length $(t_i - t_{i-1}) \wedge (t_{i+1} - t_i)$ that contain t_i do not contain any of the other points. \square

Lemma 28 *There exists a universal constant $\alpha_1 \in \mathbb{R}$ and a function $\alpha_2 : \mathbb{N} \rightarrow \mathbb{R}$ such that for any $k \geq 1$ and $i \in \{1, \dots, k\}$*

$$\mathbb{E}[\log[(Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i)] | N_X = k] = \alpha_1 - \alpha_2(k).$$

Proof: Let $\tilde{Y}_1, \dots, \tilde{Y}_k$ denote the order statistics of k independent $[0, 1[$ -uniformly distributed random variables, and let (\bar{Y}_i) denote the random jump positions of a Poisson process of intensity 1 on $[0, \infty[$. First let $i \in \{1, \dots, k-1\}$

$$\begin{aligned} \mathbb{E}[\log(Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i) | N_X = k] &= \mathbb{E} \log \left((\tilde{Y}_i - \tilde{Y}_{i-1}) \wedge (\tilde{Y}_{i+1} - \tilde{Y}_i) \right) \\ &= \mathbb{E} \log \left(\frac{\bar{Y}_i - \bar{Y}_{i-1}}{\bar{Y}_{k+1}} \wedge \frac{\bar{Y}_{i+1} - \bar{Y}_i}{\bar{Y}_{k+1}} \right) \\ &= \mathbb{E} \log \left((\bar{Y}_1 - \bar{Y}_0) \wedge (\bar{Y}_2 - \bar{Y}_1) \right) - \mathbb{E}[\log \bar{Y}_{k+1}]. \end{aligned}$$

For the second equality see e.g. [5], Proposition 13.15.

Setting $\alpha_1 := \mathbb{E} \log((\bar{Y}_1 - \bar{Y}_0) \wedge (\bar{Y}_2 - \bar{Y}_1))$ and $\alpha_2(k) := \mathbb{E}[\log \bar{Y}_{k+1}]$ finishes the proof in this case. The statement follows analogously for $i = k$. \square

Furthermore, we will need asymptotic estimates for

$$A := N_X \left(\prod_{i=1}^{N_X} (Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i) \right)^{1/N_X}$$

and

$$R_\beta := N_X \log_+ \beta \left(\prod_{i=1}^{N_X} (Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i) \right)^{1/N_X} = \log_+ \prod_{i=1}^{N_X} \beta (Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i),$$

where $\beta > 0$.

Lemma 29 *One has*

$$\mathbb{E}R_\beta \geq \lambda \log \beta + c$$

for the constant $c = c(\lambda) = \lambda \alpha_1 - \mathbb{E}[N_X \alpha_2(N_X)] \in \mathbb{R}$, where α_1 and α_2 are as in the previous lemma. Moreover,

$$\lim_{\beta \rightarrow \infty} \beta \mathbb{E} \left(\frac{N_X}{\beta} \wedge A \right) = \lambda.$$

Proof: Applying Lemma 28 we get

$$\begin{aligned} \mathbb{E}R_\beta &= \mathbb{E}[\mathbb{E}[R_\beta | N_X]] \geq \mathbb{E} \sum_{i=1}^{N_X} \left[\log \beta + \mathbb{E}[\log(Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i) | N_X] \right] \\ &= \mathbb{E}N_X \left[\log \beta + \alpha_1 - \alpha_2(N_X) \right] = \lambda \log \beta + c. \end{aligned}$$

The second statement is an immediate consequence of the monotone convergence theorem: since $A > 0$ a.s. one has

$$\beta \mathbb{E} \left[\frac{N_X}{\beta} \wedge A \right] = \mathbb{E}[N_X \wedge \beta A] \rightarrow \mathbb{E}N_X = \lambda.$$

□

We are now in the position to prove Theorem 24.

Proof of Theorem 24: Let \hat{X} be $\mathbf{D}([0, 1[, E)$ -valued reconstruction with $I(X; \hat{X}) \leq r$ for some fixed $r \geq 0$. Furthermore, we denote by

$$\begin{aligned} G(k, (I_1, \dots, I_k), (a_0, \dots, a_k)) \\ = I(X; \hat{X} | N_X = k, \pi_k(Y) = (I_1, \dots, I_k), (A_0, \dots, A_k) = (a_0, \dots, a_k)) \end{aligned}$$

the conditional mutual information of X and \hat{X} given N_X , $\pi_{N_X}(Y)$, and (A_0, \dots, A_{N_X}) . We consider the non-negative random variable $R = G(N_X, \pi_{N_X}(Y), (A_0, \dots, A_k))$.

Since $(N_X, \pi_{N_X}(X), (A_0, \dots, A_k))$ is $\sigma(X)$ -measurable one has

$$r \geq I(X; \hat{X}) \geq I(X; \hat{X} | N_X, \pi_{N_X}(Y), (A_0, \dots, A_k)) = \mathbb{E}R.$$

Moreover, Lemma 26 together with Lemma 27 implies that

$$\mathbb{E}\rho_{\mathbf{D}}(X, \hat{X}) \geq \frac{\varepsilon_0}{4e} \mathbb{E} \left[N_X \left(\prod_{i=1}^{N_X} (Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i) \right)^{1/N_X} e^{-R/N_X} \right].$$

In order to get a lower bound for the coding error we next analyze the minimization problem

$$\mathbb{E} \left[N_X \left(\prod_{i=1}^{N_X} (Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i) \right)^{1/N_X} e^{-\bar{R}/N_X} \right] = \min!$$

where the infimum is taken over all non-negative random variables \bar{R} satisfying $\mathbb{E}\bar{R} \leq r$.

We let again $A = N_X (\prod_{i=1}^{N_X} (Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i))^{1/N_X}$. Using Lagrange multipliers one gets that for every $\beta > 0$

$$R_\beta = N_X \log_+ \frac{\beta A}{N_X} = N_X \log_+ \beta \left(\prod_{i=1}^{N_X} (Y_i - Y_{i-1}) \wedge (Y_{i+1} - Y_i) \right)^{1/N_X},$$

is a minimizer when $r = r_\beta := \mathbb{E}R_\beta = \mathbb{E}[N_X \log_+ \frac{\beta A}{N_X}]$. Moreover, elementary computations

give that the corresponding minimal value in the minimization problem is

$$d_\beta := \mathbb{E} \left[A \exp \left(-\log_+ \frac{\beta A}{N_X} \right) \right] = \mathbb{E} \left(\frac{N_X}{\beta} \wedge A \right).$$

For given $r \geq 0$ we now choose $\beta = \beta(r) = \exp((r - c)/\lambda)$ where c is as in Lemma 29. Then $r = \lambda \log \beta + c \leq \mathbb{E} R_\beta$ and due to the variational formula above one has

$$D(r | X, \rho_{\mathbf{D}}, 1) \geq \frac{\varepsilon_0}{4e} \mathbb{E} \left(\frac{N_X}{\beta(r)} \wedge A \right).$$

Thus letting r tend to infinity we get

$$D(r | X, \rho_{\mathbf{D}}, 1) \gtrsim \frac{\varepsilon_0}{4e} \frac{\lambda}{\beta(r)} = \frac{\varepsilon_0}{4e} \lambda \exp(-(r - c)/\lambda).$$

Thus, one has for all sufficiently large r that

$$D(r | X, \rho_{\mathbf{D}}, 1) \geq \frac{1}{8e} \lambda e^{-c/\lambda} \varepsilon_0 e^{-r/\lambda} = C_\lambda \varepsilon_0 e^{-r/\lambda},$$

where

$$C_\lambda = \frac{\lambda}{8e} e^{c/\lambda} \text{ and } c = \lambda \alpha_1 - \mathbb{E}[N_X \alpha_2(N_X)].$$

Moreover, α_1 and α_2 can be expressed in terms of i.i.d. standard exponential random variables (e_i) as $\alpha_1 = \mathbb{E} \log(e_1 \wedge e_2)$ and $\alpha_2(n) = \mathbb{E} \log \sum_{i=1}^{n+1} e_i$, cf. the proof of Lemma 28.

After some calculations (using Mathematica) one obtains

$$\alpha_1 = \int_0^\infty (\log x) 2e^{-2x} dx = -\gamma - \log 2, \quad \alpha_2(n) = \frac{\Gamma'(n+1)}{\Gamma(n+1)},$$

where $\gamma = 0.57721\dots$ is the Euler–Mascheroni constant and Γ is the Gamma function. Some more calculations show that

$$c/\lambda = -\gamma - \log 2 - \log \lambda - \int_\lambda^\infty x^{-1} e^{-x} dx - \frac{1 - e^{-\lambda}}{\lambda}.$$

Closer analysis of this term shows that

$$\lim_{\lambda \rightarrow \infty} C_\lambda = \frac{1}{8e} e^{-\gamma}, \quad \lim_{\lambda \rightarrow 0} C_\lambda/\lambda = \frac{1}{8e^2},$$

which altogether shows that C_λ can be estimated from below by $D \min(1, \lambda)$ with some absolute constant $D > 0$. \square

Acknowledgements. The research of Frank Aurzada was supported by the DFG Research Center MATHEON “Mathematics for key technologies” in Berlin. Christian Vormoor was supported by the DFG Graduiertenkolleg 251.

References

- [1] S. Ankirchner, S. Dereich, and P. Imkeller, The Shannon information of filtrations and the additional logarithmic utility of insiders, *Ann. Probab.* 34 (2006) 743-778.
- [2] F. Aurzada and S. Dereich, The coding complexity of Lévy processes, Preprint (2007), available from: <http://arxiv.org/abs/0707.3040>
- [3] J. Bertoin, Lévy processes, volume 121 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, UK, 1996.
- [4] N. H. Bingham, C. M. Goldie, and J. L. Teugels, Regular Variation, volume 27 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, UK, 1989.
- [5] L. Breiman, Probability, volume 7 of Classics in Applied Mathematics, SIAM, Philadelphia, USA, 1993.
- [6] T. M. Cover and J. A. Thomas, Elements of information theory, Wiley Series in Telecommunications, John Wiley & Sons, Inc., New York, USA, 1991.
- [7] J. Creutzig, S. Dereich, Th. Müller-Gronbach, and K. Ritter, Infinite-dimensional quadrature and approximation of distributions, Preprint (2007).
- [8] I. Csiszar, I -divergence geometry of probability distributions and minimization problems, *Ann. Probab.* 3 (1975) 146–158.
- [9] S. Dereich, The coding complexity of diffusion processes under supremum norm distortion, to appear in: *Stochastic Process. Appl.*, available from: <http://dx.doi.org/10.1016/j.spa.2007.07.003>, 2007.
- [10] S. Dereich, The coding complexity of diffusion processes under $L^p[0, 1]$ -norm distortion, to appear in: *Stochastic Process. Appl.*, available from: <http://dx.doi.org/10.1016/j.spa.2007.07.002>, 2007.
- [11] 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 (2003) 249–265.
- [12] S. Dereich and M. Scheutzow, High resolution quantization and entropy coding for fractional Brownian motion, *Electronic J. Prob.* 11 (2006) 700–722.
- [13] S. Graf and H. Luschgy, Foundations of Quantization for Probability Distributions, volume 1730 of Lecture Notes in Mathematics, Springer, Berlin, Germany, 2000.
- [14] S. Ihara, Information theory for continuous systems, World Scientific, Singapore, 1993.
- [15] A. N. Kolmogorov, Three approaches to the quantitative definition of information, *Internat. J. Comput. Math.* 2 (1968) 157–168.

- [16] H. Luschgy and G. Pagès, Sharp asymptotics of the functional quantization problem for Gaussian processes, *Ann. Probab.* 32 (2004) 1574–1599.
- [17] H. Luschgy and G. Pagès, Functional quantization of a class of Brownian diffusions: a constructive approach, *Stochastic Process. Appl.* 116 (2006) 310–336.
- [18] H. Luschgy and G. Pagès, Functional quantization rate and mean pathwise regularity of processes with an application to Lévy processes, to appear in: *Ann. Appl. Probab.* 2007.
- [19] G. Pagès and J. Printems, Functional quantization for pricing derivatives, Université de Paris VI, LPMA no. 930, Preprint, 2004.
- [20] C. Vormoor, High resolution coding of point processes and the Boolean model, PhD thesis, Technische Universität Berlin, 2007.