Risk Bounds for Lévy Processes in the PAC-Learning Framework

Chao Zhang
School of Computer Engineering
Nanyang Technological University

Dacheng Tao
School of Computer Engineering
Nanyang Technological University

Abstract

Lévy processes play an important role in the stochastic process theory. However, since samples are non-i.i.d., statistical learning results based on the i.i.d. scenarios cannot be utilized to study the risk bounds for Lévy processes. In this paper, we present risk bounds for non-i.i.d. samples drawn from Lévy processes in the PAC-learning framework. In particular, by using a concentration inequality for infinitely divisible distributions, we first prove that the function of risk error is Lipschitz continuous with a high probability, and then by using a specific concentration inequality for Lévy processes, we obtain the risk bounds for non-i.i.d. samples drawn from Lévy processes without Gaussian components. Based on the resulted risk bounds, we analyze the factors that affect the convergence of the risk bounds and then prove the convergence.

1 Introduction

Most existing results of statistical learning theory, *e.g.,* probably approximately correct (PAC) learning framework (Valiant, 1984), are based on some classical statistical results (*e.g.*, the central limit theorem), which are valid for the independent and identically distributed (i.i.d.) samples.[1] One of the major concerns in statistical learning theory is the risk bound. Risk bounds of learning algorithms measure the probability that a function produced by an algorithm has a sufficiently small error. Vapnik (1999) gave asymptotic risk bounds in terms of complexity measures of function classes such as VC dimensions. Various risk bounds can be obtained by using different concentration inequalities and symmetrization inequalities. Bartlett *et al.* (2002) introduced local Rademacher complexities and presented sharp risk bounds for a particular function class. Sridharan *et al.* (2008) showed that the empirical minimizer can converge to the optimal value at rate $1/N$ if the empirical minimization of a stochastic objective is λ-strongly convex and the stochastic component is linear. Tsybakov (2004) presented a classifier that automatically adapts to the margin condition and its error bound approaches to *zero* at a fast rate $1/N$. In a hypothesis with some noise conditions, the rates of error bounds are always faster than $1/\sqrt{N}$ (Bousquet, 2002). However, these results share an identical assumption that the samples are i.i.d., and they are no longer valid for non-i.i.d. samples.

In practical applications, the i.i.d. assumption for samples is not always valid. For example, some financial and physical behaviors are temporally dependent and thus the aforementioned research results are unsuitable. Mohri and Rostamizadeh (2008) gave the Rademacher complexity-based risk bounds for stationary β-mixing sequences. They can be deemed as transitions between i.i.d. scenarios and non-i.i.d. scenarios, where the dependence between samples diminishes along time. Especially, by utilizing a technique of independent blocks (Yu, 1994), the samples drawn from a β-mixing sequence can be transformed to an i.i.d. case and some classical results for i.i.d. cases can be applied to obtain the risk bounds. Moreover, there are also some works about the uniform laws for dependent processes (Nobel and Dembo, 1993).

In this paper, we focus on risk bounds for Lévy processes that are the stochastic processes with stationary and independent increments.

Definition 1.1 *A stochastic process* $(Z_t)_{t\geq 0}$ *on* R^d *is a Lévy process if it satisfies the following conditions:*

1. $Z_0 = 0$, *a.s.*

[1] Note that Kolmogorov's central limit theorem is valid for identically distributed samples (Petrov, 1995).

Appearing in Proceedings of the 13^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.

2. *For any $n \geq 1$ and $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$ the random variables*

$$Z_{t_0}, Z_{t_1} - Z_{t_0}, \cdots, Z_{t_n} - Z_{t_{n-1}}$$

are independent.

3. *The increments are stationary, i.e., the distribution of $Z_{s+t} - Z_s$ is independent of s.*

4. *The process is right continuous, i.e., for any $s \geq t \geq 0$ and $\epsilon > 0$, we have*

$$\lim_{s \to t} \mathrm{P}\Big[|Z_t - Z_s| > \epsilon\Big] = 0.$$

Lévy processes play an important role in financial mathematics, physics, engineering and actuarial science (Barndorff-Nielsen *et al.*, 2001; Applebaum, 2004a). Lévy processes contain many important special examples, *e.g.*, Brownian motion, Poisson processes, stable processes and subordinators. Lévy processes have been regarded as prototypes of semimartingales and Feller-Markov processes (Applebaum, 2004b; Sato, 2004). Figueroa-López and Houdré (2006) used projection estimators to estimate the Lévy density, and then gave a bound to exhibit the discrepancy between a projection estimator and the orthogonal projection by using the concentration inequalities for functionals of Poisson integrals. This paper is focused on the risk bounds for Lévy processes in the PAC-learning framework.

The PAC-learning framework was proposed by Valiant (1984) and intends to obtain, with high probability, a hypothesis that is a good approximation to an unknown target by successful learning. In this framework, a learner receives some samples and then selects an appropriate function from a function class based on these samples. The selected function has a low risk error with a high probability. Given an input space $\mathcal{X} \in \mathrm{R}^I$ and its corresponding output space $\mathcal{Y} \in \mathcal{R}^J$, formally, we define $\mathcal{Z} = (\mathcal{X}, \mathcal{Y}) \subset \mathrm{R}^{I \times J}$ and assume that $\mathcal{Z} = \{Z_t\}_{t \geq 0}$ is a Lévy process without Gaussian components with $Z_t = (\mathbf{x}_t, \mathbf{y}_t)$.

We expect to find a function $T : \mathcal{X} \to \mathcal{Y}$ that, given a new input $\mathbf{x}_t$ ($t > 0$), accurately predicts the output $\mathbf{y}_t$. Particularly, for a loss function $\ell : \mathcal{Y}^2 \to \mathrm{R}$, the target function T minimizes the expected risk, for any $t > 0$,

$$\mathrm{E}_t(\ell(T(\mathbf{x}_t), \mathbf{y}_t)) = \int \ell(T(\mathbf{x}_t), \mathbf{y}_t) dP_t, \qquad (1)$$

where P_t stands for the distribution of $Z_t = (\mathbf{x}_t, \mathbf{y_t})$. Since P_t is unknown, T usually cannot be directly obtained by using (1). Therefore, given a function class $\mathcal{G}$ and a sample set $\{Z_{t_n}\}_{n=1}^N \subset \mathcal{Z}$ with $Z_{t_n} = (\mathbf{x}_{t_n}, \mathbf{y}_{t_n})$, the estimation to the target function T is determined by minimizing the following empirical risk

$$E_N(\ell(g(\mathbf{x}), \mathbf{y})) = \frac{1}{N} \sum_{n=1}^{N} \ell(g(\mathbf{x}_{t_n}), \mathbf{y}_{t_n}), \;\; g \in \mathcal{G}, \quad (2)$$

which is considered as an approximation to the expected risk (1). Next, we define the loss function class

$$\mathcal{F} = \{Z \mapsto \ell(g(\mathbf{x}), \mathbf{y}) : g \in \mathcal{G}\}.$$

and call $\mathcal{F}$ the function class in the rest of the paper. For any $t > 0$ and a sample set $\{Z_{t_n}\}_{n=1}^N \in \mathcal{Z}$ ($t_1 < t_2 < \cdots < t_N$), we shortly denote, for any $f \in \mathcal{F}$,

$$\mathrm{E}_t f(Z_t) = \int f(Z_t) dP_t \;, \;\; \mathrm{E}_N f = \frac{1}{N} \sum_{n=1}^{N} f(Z_{t_n}),$$

and define the risk error function with respect to Z_t as

$$\Phi(Z_t) = \sup_{f \in \mathcal{F}} \Big| \mathrm{E}_N f - \mathrm{E}_t f(Z_t) \Big|.$$

Similar to the processes of obtaining risk bounds for i.i.d. samples, *e.g.*, (Bartlett *et al.*, 2002), we use a specific concentration inequality to obtain risk bounds for the empirical processes of Lévy processes. Houdré and Marchal (2008) discussed median, concentration and fluctuations for Lévy processes and gave a concentration inequality for Lévy processes without Gaussian components under the 1-Lipschitz continuity. According to the discussions in (Houdré and Marchal, 2008) and (Houdré, 2002), the concentration inequality also holds for the λ-Lipschitz continuity ($\lambda > 0$). By using the new concentration inequality, we achieve an upper bound of risk error $\Phi(Z_t)$ under the PAC-learning framework, wherein Z_t is drawn from an unknown Lévy process $\mathcal{Z}$ without Gaussian components. Its characteristic exponent $\psi_1(\theta)$ of $Z_1 \in \mathcal{Z}$ is given, for all $\theta \in \mathrm{R}^{I \times J}$, by

$$\begin{aligned}\psi_{\mathcal{Z}}(\theta) =& i < \theta, \mathbf{a} > + \int_{\mathrm{R}^{I \times J}} (\mathrm{e}^{i<\theta, y>} \\ & - 1 - i < \theta, y > \mathbf{1}_{\|y\| \leq 1}) \nu(dy), \end{aligned} \qquad (3)$$

where $\mathbf{a} \in \mathrm{R}^{I \times J}$ and $\nu \not\equiv 0$ is a Lévy measure. Moreover, in (3), $< \cdot, \cdot >$ and $\| \cdot \|$ are the Euclidean inner product and a norm in $\mathrm{R}^{I \times J}$, respectively. Then, for any $t > 0$, we have

$$\mathrm{E}_t \exp(i < \theta, Z_t >) = \exp\left(t \psi_1(\theta)\right). \qquad (4)$$

The above (4) implies the process $\mathcal{Z}$ is completely determined by the distribution of Z_1 and $(\mathbf{a}, 0, \nu)$ is the generating triplet of $\mathcal{Z}$. In the next section, we provide details of (3) and (4).

The rest of this paper is organized as follows. Section 2 introduces Lévy processes. The main results are presented in Section 3. Section 4 gives the proofs of some lemmas and theorems and Section 5 concludes the paper.

2 Lévy Processes

In this section, we briefly introduce Lévy processes. Details are given in (Applebaum, 2004b; Sato, 2004).

We first introduce the infinitely divisible distributions which are strongly related to Lévy processes. Afterward, we show that for a Lévy process $\{Z_t\}_{t\geq 0}$, if $Z_t \in \{Z_t\}_{t\geq 0}$ $(t > 0)$, the distribution of Z_t is infinitely divisible.

Definition 2.1 *A real-valued random variable Z has an infinitely divisible distribution if for any $N = 1, 2, \cdots$, there is a sequence of i.i.d. random variables $\{Z_n^{(N)}\}_{n=1}^N$ such that $Z \sim Z_1^{(N)} + \cdots + Z_N^{(N)}$ where $\sim$ is equality in distribution.*

According to the definition, if a random variable has an infinitely divisible distribution, it can be separated into the sum of arbitrary number of i.i.d. random variables. Next, we introduce the characteristic exponent of an infinitely divisible distribution (Sato, 2004). Before the statement, we need the following definition (Applebaum, 2004b).

Definition 2.2 *Let ν be a Borel measure defined on $\mathrm{R}^d\backslash\{0\}$. This ν will be a Lévy measure if*

$$\int_{\mathrm{R}^d\backslash\{0\}} \min\{\|y\|^2, 1\}\nu(dy) < \infty, \tag{5}$$

and $\nu(\{0\}) = 0$.

The Lévy measure describes the expected number of a certain height jump in a time interval of unit length 1. Then, the characteristic exponent of an infinitely divisible random variable is shown in the following theorem.

Theorem 2.3 (Lévy-Khintchine) *A Borel probability measure μ of a random variable $Z \in \mathrm{R}^d$ is infinitely divisible if there exists a triple $(\mathbf{a}, \mathbf{A}, \nu)$, where $\mathbf{a} \in \mathrm{R}^d$, a positive-definite symmetric $d \times d$ matrix $\mathbf{A}$ and a Lévy measure ν on $\mathrm{R}^d\backslash\{0\}$ such that, for all $\theta \in \mathrm{R}^d$, the characteristic exponent ψ_μ is of the form*

$$\psi_\mu(\theta) = i < \mathbf{a}, \theta > - \frac{1}{2} < \theta, \mathbf{A}\theta > + \int_{\mathrm{R}^d\backslash\{0\}} \Big[\mathrm{e}^{i<\theta,y>} - 1 - i < \theta, y > \mathbf{1}_{\|y\|\leq 1} \Big] \nu(dy). \tag{6}$$

As stated above, an infinitely divisible distribution can be completely determined by a triple $(\mathbf{a}, \mathbf{A}, \nu)$ wherein "$\mathbf{a}$" stands for the drift of a Brownian motion, "$\mathbf{A}$" is a Gaussian component and "ν" is a Lévy measure.

Next, we show that for any Lévy process, there must be an infinitely divisible random variable corresponding to it, and vice versa. It is to be pointed out that for any Lévy process $\{Z_t\}_{t\geq 0}$, the distributions of Z_t $(t > 0)$ are all determined by the distribution of Z_1. The issue has been sketched in (Applebaum, 2004b).

Lemma 2.4 *Let $\{Z_t\}_{t\geq 0}$ be a Lévy process.*

(i) For any $t > 0$, Z_t has an infinitely divisible distribution.

(ii) For any $t > 0$ and $\theta \in \mathrm{R}^d$, let the characteristic exponent

$$\psi_t(\theta) = \log(\mathrm{E}_t(\mathrm{e}^{i\theta t})). \tag{7}$$

Then we have

$$\psi_t(\theta) = t\psi_1(\theta), \tag{8}$$

where $\psi_1(t)$ is the characteristic exponent of Z_1.

The above lemma shows that any Lévy process has the property that for all $t \geq 0$,

$$\mathrm{E}_t(\mathrm{e}^{i\theta Z_t}) = \mathrm{e}^{t\psi_1(\theta)}, \tag{9}$$

where $\psi_1(t)$ is the characteristic exponent of Z_1 that has an infinitely divisible distribution, *i.e.,* any Lévy process corresponds to an infinitely divisible distribution. The next theorem shows (Sato, 2004) that, given an infinitely divisible distribution, one can construct a Lévy process $\{Z_t\}_{t\geq 0}$ such that Z_1 has that distribution.

Theorem 2.5 *Suppose that $\mathbf{a} \in \mathrm{R}^d$, $\mathbf{A}$ is a positive-definite symmetric $d \times d$ matrix and ν is a measure concentrated on $\mathrm{R}\backslash\{0\}$ such that $\int(1 \wedge \|y\|^2)\nu(dy) < \infty$. By the triple $(\mathbf{a}, \mathbf{A}, \nu)$, for each $\theta \in \mathrm{R}^d$, define*

$$\psi(\theta) = i < \mathbf{a}, \theta > - \frac{1}{2} < \theta, \mathbf{A}\theta > + \int \Big(\mathrm{e}^{i<\theta,y>} - 1 - i < \theta, y > \mathbf{1}_{\|y\|<1}\nu(dy) \Big). \tag{10}$$

Then, there exists a Lévy process $\{Z_t\}_{t\geq 0}$ where Z_1 has the characteristic exponent in the form of (10).

Note that according to Lemma 2.4 and Theorem 2.5, a Lévy process $\{Z_t\}_{t\geq 0}$ can be distinguished by the triple $(\mathbf{a}, \mathbf{A}, \nu)$ of the Z_1's distribution. Thus, we call $(\mathbf{a}, \mathbf{A}, \nu)$ the generating triplet of $\{Z_t\}_{t\geq 0}$. Furthermore, according to Lemma 2.4 and Theorem 2.5, we can obtain the following result (Sato, 2004), which is necessary for providing the main results.

Lemma 2.6 *Let $(\mathbf{a}, \mathbf{A}, \nu)$ be the generating triplet of a Lévy process $\{Z_t\}_{t\geq 0}$. Then, at any time $t > 0$, $Z_t \in \{Z_t\}_{t\geq 0}$ has an infinitely divisible distribution with the triplet $(\mathbf{a}t, \mathbf{A}t, \nu t)$.*

This paper mainly concerns the Lévy process with the generating triplet $(\mathbf{a}, 0, \nu)$ wherein the Gaussian component is *zero*, as shown in (3).[2]

[2] We refer to the reference (Sato, 2004) for the detailed discussion about the effect of a triple $(\mathbf{a}, \mathbf{A}, \nu)$ to the path of the corresponding Lévy process.

3 Main Results

Given a Lévy process $\{Z_t\}_{t\geq 0}$ with the Lévy measure ν, we give the following definitions (Marcus and Rosiński, 2001). For any $C > 0$, define

$$V(C) = \int_{\|y\|\leq C} \|y\|^2 \nu(dy), \tag{11}$$

and

$$M(C) = \int_{\|y\|\geq C} \|y\| \nu(dy). \tag{12}$$

Next, let $\overline{\nu}$ be the tail of ν, *i.e.*,

$$\overline{\nu}(C) = \int_{\|y\|>C} \nu(dy). \tag{13}$$

Moreover, given a function class $\mathcal{F}$ and a Lévy process $\{Z_t\}_{t\geq 0}$ with characteristic exponent (3), for any $t \geq 0$, let

$$\Phi(Z_t) = \sup_{f\in\mathcal{F}} \Big| \mathrm{E}_N f - \mathrm{E}_t f(Z_t) \Big|, \tag{14}$$

and let f_t be the solution to the supremum of $\Phi(Z_t)$ over the function class $\mathcal{F}$. Then, let $\mathcal{F}^* = \{f_t\}_{t\geq 0}$ be the set of all f_t $(t \geq 0)$.

In this paper, we need the following mild assumptions:

$(A1)$ The Lévy process $\mathcal{Z} = (Z_t)_{t\geq 0}$ has a finite expectation and is centered, *i.e.*, for any $t \geq 0$,

$$\mathrm{E}_t Z_t = 0. \tag{15}$$

$(A2)$ There exists a constant K such that for every $C > 0$, we have

$$M(C) \leq K \frac{V(C)}{C}. \tag{16}$$

$(A3)$ There exists a constant $A > 0$ such that, for any $C > 0$, we have

$$\overline{\nu}(C) \leq A \frac{V(C)}{C^2}. \tag{17}$$

$(A4)$ For any $s, t > 0$ and $s \neq t$, there exists a positive constant α such that

$$|f_t(Z_t) - f_s(Z_s)| \leq \alpha \|Z_t - Z_s\|, \tag{18}$$

where $Z_s, Z_t \in \mathcal{Z}$ and $f_s, f_t \in \mathcal{F}^*$.

$(A5)$ There exists a positive constant β such that for any $f \in \mathcal{F}$ and $Z', Z'' \in \mathrm{R}^{I\times J}$, we have

$$|f(Z') - f(Z'')| \leq \beta \|Z' - Z''\|. \tag{19}$$

Remark:

(i) If $\mathcal{Z}$ has non-centered finite mean, one can consider the Lévy process $\{Z_t - \mathrm{E}Z_t\}_{t\geq 0}$, where the assumption $(A1)$ holds.

(ii) The assumption $(A5)$ implies that elements of $\mathcal{F}$ satisfy the β-Lipschitz continuity.

Subsequently, we begin the main discussion of this paper. First, we prove that the function $\Phi(Z_t)$ (cf. (14)), with a high probability, is Lipschitz continuous with respect to Z_t. Afterward, according to a concentration inequality for Lévy process (Houdré and Marchal, 2008), we obtain the risk bound of $\Phi(Z_t)$ in the PAC-learning framework. Then, we analyze the terms in the resulting bound and prove the convergence.

3.1 Risk Bounds

Under the above assumptions, we first consider the Lipschitz continuity of the function $\Phi(Z_t)$.

Lemma 3.1 *Given a Lévy process $\mathcal{Z} = \{Z_t\}_{t\geq 0}$ with characteristic exponent (3), $\{Z_{t_n}\}_{n=1}^N$ is an ordered sample set drawn from $\mathcal{Z}$. If assumptions $(A4) - (A5)$ are valid, then we have that, for any $\eta > 0$ and $Z_s, Z_t \in \mathcal{Z}$ $(s \neq t)$, with the probability at least $1 - \delta_t - \delta_s$,*

$$|\Phi(Z_t) - \Phi(Z_s)| \leq (2\alpha + 6\eta) \|Z_t - Z_s\|, \tag{20}$$

where

$$\delta_t = \exp\left\{ \frac{\eta\|Z_t - Z_s\|}{\beta S} - \left(\frac{\eta\|Z_t - Z_s\|}{\beta S} + \frac{tV}{S^2} \right) \times \log\left(1 + \frac{\eta\|Z_t - Z_s\| S}{t\beta V} \right) \right\}, \tag{21}$$

and

$$\delta_s = \exp\left\{ -\frac{\eta\|Z_t - Z_s\|}{\beta S} + \left(\frac{\eta\|Z_t - Z_s\|}{\beta S} - \frac{sV}{S^2} \right) \times \log\left(1 - \frac{\eta\|Z_t - Z_s\| S}{s\beta V} \right) \right\} \tag{22}$$

with $S = \inf\{\rho > 0 : \nu(y : \|y\| > \rho) = 0\}$.

The above lemma shows that for a given sample set $\{Z_{t_n}\}_{n=1}^N$, $\Phi(Z_t)$ is Lipschitz continuous with a high probability.

For some positive real number c, let

$$h_c(t) = \inf\left\{ C > 0 : \frac{V(C)}{C^2} = \frac{c}{t} \right\}, \quad t > 0. \tag{23}$$

Then, we can arrive at the main theorem of this paper.

Theorem 3.2 *Given a function class $\mathcal{F}$ and a sample set $\{Z_{t_n}\}_{n=1}^N$ drawn from a Lévy process with characteristic exponent* (3). *If the assumptions $(A1)-(A5)$ are valid and there exists a constant λ such that $(2\alpha+6\eta)\leq\lambda$, then with probability at least $\delta_L-\delta$, the following bound holds*

$$\Phi(Z_t) \leq (b+\lambda cK)h_c(t) + \frac{\beta}{N}\sum_{n=1}^{N} \mathrm{E}_{|t-t_n|}\left\|Z_{|t-t_n|}\right\|, \quad (24)$$

where

$$\delta = Ac + \exp\left\{\frac{b}{\lambda} - (\frac{b}{\lambda}+c)\log\left(1+\frac{b}{\lambda c}\right)\right\}, \quad (25)$$

and δ_L is the probability that makes Lemma 3.1 hold.

The above theorem gives a bound of $\Phi(Z_t)$. In Lemma 3.1, we have proven that the Lipschitz continuity of $\Phi(Z_t)$ holds with a high probability. Thus, Theorem 3.2 is valid with a high probability.

3.2 Convergence Analysis

We analyze the terms at the right hand side of (29) and discuss some factors that affect the convergence. The following lemma gives a bound to the fluctuations of Z_t $(t>0)$ (Houdré and Marchal, 2008).

Lemma 3.3 *Let $C_0(t)$ $(t>0)$ be the solution in C of the equation:*

$$\frac{V(C)}{C^2} + \frac{M(C)}{C} = \frac{1}{t}. \quad (26)$$

If the assumption $(A1)$ is valid, then

$$\frac{1}{4}C_0(t) \leq \mathrm{E}_t\|Z_t\| \leq \frac{17}{8}C_0(t), \quad (27)$$

and the factor 17/8 can be replaced by 5/4 when $\{Z_t\}_{t\geq 0}$ is symmetric.

Furthermore, if assumptions $(A1)-(A2)$ are both valid, then we have

$$h_{\frac{1}{1+K}}(t) \leq C_0(t) \leq h_1(t). \quad (28)$$

The above lemma shows that if assumptions $(A1)-(A2)$ are valid, we can use $h_c(t)$ to bound $\mathrm{E}\|Z_t\|$. Therefore, by combining Theorem 3.2 and Lemma 3.3, we have the following result.

Corollary 3.4 *Given a function class $\mathcal{F}$ and a sample set $\{Z_{t_n}\}_{n=1}^N$ drawn from a Lévy process with characteristic exponent* (3). *If assumptions $(A1)-(A5)$ are valid and there exists a constant λ such that $(2\alpha+6\eta)\leq\lambda$, then with probability at least $\delta_L-\delta$, the following bound holds*

$$\Phi(Z_t) \leq (b+\lambda cK)h_c(t) + \frac{17\beta}{8N}\sum_{n=1}^{N} h_1(|t-t_n|), \quad (29)$$

where

$$\delta = Ac + \exp\left\{\frac{b}{\lambda} - (\frac{b}{\lambda}+c)\log\left(1+\frac{b}{\lambda c}\right)\right\}, \quad (30)$$

δ_L is the probability that makes Lemma 3.1 hold, and the factor 17/8 can be replaced by 5/4 when $\{Z_t\}_{t\geq0}$ is symmetric.

The terms at the right hand side of (29) affect the convergence of the risk bound with respect to N (the number of samples). Next, we present a convergence theorem about the risk bound for Lévy processes.

Theorem 3.5 *Follow the notations and conditions of Theorem 3.2 and Corollary 3.4, and assume that $0<c<<1/A$, $c=o(1)$ and $0<b=o(\sqrt{c})$ as $N\to\infty$, wherein $o(\cdot)$ stands for the infinitesimal of higher order. If $\sum_{n=1}^N h_1(|t-t_n|)=o(N)$ and there exists a constant C_ν such that*

$$\int_{\mathrm{R}^{I\times J}} \|y\|^2\nu(dy) < C_\nu, \quad (31)$$

then we have that for any given $t>0$, with probability at least $\delta_L-\delta$,

$$\lim_{N\to 0}\Phi(Z_t) = 0. \quad (32)$$

On the one hand, from the proof of Theorem 3.5, if there exists a positive constant C_ν satisfying (31), for any $t>0$, the rate of $(b+\lambda cK)h_c(t)\to 0$ $(N\to\infty)$ is completely determined by the selection of b, c. On the other hand, according to (30), the probability δ can be deemed as a function of b and c, and thus the values of b and c must satisfy the constraint $\delta<1$. According to (30), it is clear that if $b,c>0$ and $c<<1/A$, the constraint $\delta<1$ holds. Therefore, in Theorem 3.5, $(b+\lambda cK)h_c(t)$ can reach *zero* at an arbitrary rate and the convergence is mainly affected by the term $(1/N)\sum_{n=1}^N h_1(|t-t_n|)$.

Recalling the definition of $h_1(t)$ (*cf.* (23)), we have that for a given $t>0$, the convergence of $\Phi(Z_t)$ is determined by the Lévy measure ν. According to the definition of Lévy measure ν (*cf.* (5)), if many large jumps frequently appear in the path of a Lévy process, the term $(1/N)\sum_{n=1}^N h_1(|t-t_n|)$ may be infinite when N approaches to *infinity* and thus the convergence is not valid.

4 Proofs of Lemmas and Theorems

In this section, we prove Lemma 3.1, Theorem 3.2 and Theorem 3.5, respectively.

4.1 Proof of Lemma 3.1

To prove this lemma, it is necessary to have the concentration inequality proposed by Houdré (2002).

Lemma 4.1 *Let ν have bounded support with*

$$S = \inf\{\rho > 0 : \nu(\{y : \|y\| > \rho\}) = 0\},$$

and let $V = \int_{\mathrm{R}^d} \|y\|^2 \nu(dy)$. If f is a λ-Lipschitz function, then for any $x \geq 0$, we have

$$\begin{aligned} &\mathrm{P}\Big(f(Z) - \mathrm{E}_Z f(Z) > x\Big) \\ &\leq \exp\left(\frac{x}{\lambda S} - \left(\frac{x}{\lambda S} + \frac{V}{S^2}\right)\log\left(1 + \frac{Sx}{\lambda V}\right)\right). \end{aligned} \tag{33}$$

Lemma 4.2 *Given a Lévy process with a generating triple $(\mathbf{a}, 0, \nu)$, let ν have bounded support with*

$$S = \inf\{\rho > 0 : \nu(\{y : \|y\| > \rho\}) = 0\},$$

and let $V = \int_{\mathrm{R}^d} \|y\|^2 \nu(dy)$. If assumptions $(A4)-(A5)$ are valid, then we have that with probability at least $1 - \delta_t - \delta_s$, for any real numbers $s, t > 0$ ($s \neq t$) and any $\eta > 0$,

$$|\mathrm{E}_t f_t(Z_t) - \mathrm{E}_t f_s(Z_s)| \leq (\alpha + 2\eta)\|Z_t - Z_s\|, \tag{34}$$

and for any $Z', Z'' \in \mathcal{Z}$, we have that

$$|f_t(Z') - f_s(Z'')| \leq (\alpha + 4\eta)\|Z_t - Z_s\|, \tag{35}$$

where

$$\begin{aligned} \delta_t = \exp\Bigg\{&\frac{\eta\|Z_t - Z_s\|}{\beta S} - \left(\frac{\eta\|Z_t - Z_s\|}{\beta S} + \frac{tV}{S^2}\right) \\ &\times \log\left(1 + \frac{\eta\|Z_t - Z_s\|S}{t\beta V}\right)\Bigg\}, \end{aligned}$$

and

$$\begin{aligned} \delta_s = \exp\Bigg\{&-\frac{\eta\|Z_t - Z_s\|}{\beta S} + \left(\frac{\eta\|Z_t - Z_s\|}{\beta S} - \frac{sV}{S^2}\right) \\ &\times \log\left(1 - \frac{\eta\|Z_t - Z_s\|S}{s\beta V}\right)\Bigg\}. \end{aligned}$$

Proof of Lemma 4.2.

According to Lemma 2.4 and Lemma 2.6, Z_t and Z_s have infinitely divisible distributions with the triples $(\mathbf{a}t, 0, \nu t)$ and $(\mathbf{a}s, 0, \nu s)$, respectively. According to Lemma 4.1, for any $x \geq 0$, we have

$$\mathrm{P}\Big(f_t(Z_t) - \mathrm{E}_t f_t(Z_t) > x\Big) \leq \delta_t^{(x)}, \tag{36}$$

and

$$\mathrm{P}\Big(f_s(Z_s) - \mathrm{E}_s f_s(Z_s) < -x\Big) \leq \delta_s^{(-x)}, \tag{37}$$

where

$$\delta_t^{(x)} = \exp\left\{\frac{x}{\beta S} - \left(\frac{x}{\beta S} + \frac{tV}{S^2}\right)\log\left(1 + \frac{xS}{t\beta V}\right)\right\}, \tag{38}$$

and

$$\delta_s^{(-x)} = \exp\left\{-\frac{x}{\beta S} + \left(\frac{x}{\beta S} - \frac{sV}{S^2}\right)\log\left(1 - \frac{xS}{s\beta V}\right)\right\}. \tag{39}$$

Then, we have that, with probability at least $1 - \delta_t^{(x)}$,

$$f_t(Z_t) - \mathrm{E}_t f_t(Z_t) \leq x, \tag{40}$$

and with probability at least $1 - \delta_s^{(-x)}$,

$$f_s(Z_s) - \mathrm{E}_s f_s(Z_s) \geq -x, \tag{41}$$

According to (40) and (41), we have that with probability at least $1 - \delta_t^{(x)} - \delta_s^{(-x)}$,

$$\Big(f_t(Z_t) - f_s(Z_s)\Big) - \Big(\mathrm{E}_t f_t(Z_t) - \mathrm{E}_s f_s(Z_s)\Big) \leq 2x. \tag{42}$$

According to the assumption $(A4)$ and (42), we have that with the probability at least $1 - \delta_t^{(x)} - \delta_s^{(-x)}$,

$$\begin{aligned} \Big|\mathrm{E}_t f_t(Z_t) - \mathrm{E}_s f_s(Z_s)\Big| \leq& \Big|f_t(Z_t) - f_s(Z_s)\Big| + 2x \\ \leq& \alpha\|Z_t - Z_s\| + 2x. \end{aligned} \tag{43}$$

Recall (42) and note that since Z_t and Z_s are the symbolic variables, it is valid to replace Z_t (*resp.* Z_s) with Z' (*resp.* Z''). Thus, according to (42), we have that with the probability at least $1 - \delta_t^{(x)} - \delta_s^{(-x)}$,

$$\begin{aligned} \Big|f_t(Z') - f_s(Z'')\Big| \leq& \Big|\mathrm{E}_t f_t(Z') - \mathrm{E}_s f_s(Z'')\Big| + 2x \\ \leq& \alpha\|Z_t - Z_s\| + 4x. \end{aligned} \tag{44}$$

Because x is an arbitrary nonnegative real number, we let $x = \eta\|Z_t - Z_s\|$, where $\eta > 0$ is an arbitrary real number. Then, by substituting $x = \eta\|Z_t - Z_s\|$ into (38), (39), (43) and (44), we complete the proof. □

Based on Lemma 4.2, we can arrive at the proof of Lemma 3.1.

Proof of Lemma 3.1.

First, according to Lemma 4.2, we have that for any $f_s, f_t \in \mathcal{F}^*$, with probability at least $1-\delta_t-\delta_s$,

$$\begin{aligned}\Big|\mathrm{E}_N f_t - \mathrm{E}_N f_s\Big| &= \Big|\frac{1}{N}\sum_{n=1}^{N} f_t(Z_{t_n}) - \frac{1}{N}\sum_{n=1}^{N} f_s(Z_{t_n})\Big| \\ \leq &\frac{1}{N}\sum_{n=1}^{N}\Big|f_t(Z_{t_n}) - f_s(Z_{t_n})\Big| \leq (\alpha+4\eta)\|Z_t - Z_s\|,\end{aligned} \tag{45}$$

where δ_t and δ_s are given in (21) and (22), respectively.

Thus, for any Z_t and Z_s, according to (14), (34) and (45), we have that with probability at least $1-\delta_t-\delta_s$,

$$\begin{aligned}&|\Phi(Z_t)-\Phi(Z_s)| \\ =&\Big|\sup_{f\in\mathcal{F}}|\mathrm{E}_N f - \mathrm{E}_t f(Z_t)| - \sup_{f\in\mathcal{F}}|\mathrm{E}_N f - \mathrm{E}_s f(Z_s)|\Big| \\ =&\Big||\mathrm{E}_N f_t - \mathrm{E}_t f_t(Z_t)| - |\mathrm{E}_N f_s - \mathrm{E}_s f_s(Z_s)|\Big| \\ \leq&\Big|\mathrm{E}_N f_t - \mathrm{E}_t f_t(Z_t) - \mathrm{E}_N f_s + \mathrm{E}_s f_s(Z_s)\Big| \\ \leq&\Big|\mathrm{E}_t f_t(Z_t) - \mathrm{E}_s f_s(Z_s)\Big| + \Big|\mathrm{E}_N f_t - \mathrm{E}_N f_s\Big| \\ \leq&(2\alpha+6\eta)\|Z_t - Z_s\|.\end{aligned} \tag{46}$$

This completes the proof. □

4.2 Proof of Theorem 3.2

To prove Theorem 3.2, it is necessary to introduce a concentration inequality proposed by Houdré and Marchal (2008).

Lemma 4.3 *Suppose that $\mathcal{Z} = \{Z_t\}_{t\geq 0}$ is a Lévy process with characteristic exponent (3) and f is a λ-Lipschitz function. If assumptions $(A1)-(A2)$ are valid, then for any $b>0$ and any $c,t>0$ such that $C=h_c(t)$ supports the assumption $(A3)$, we have*

$$\begin{aligned}&\mathrm{P}\Big(f(Z_t) - \mathrm{E}_t f(Z_t) \geq (b+\lambda c K)h_c(t)\Big) \\ &\quad\leq Ac + \exp\left\{\frac{b}{\lambda} - (\frac{b}{\lambda}+c)\log\left(1+\frac{b}{\lambda c}\right)\right\}.\end{aligned} \tag{47}$$

Based on the above concentration inequality, we prove the main theorem as follows.

Proof of Theorem 3.2.

We only consider the bound of $\mathrm{E}_{t_1\cdots t_N}\Phi(Z_t)$, and then the results can be obtained by Lemma 3.1 and Lemma 4.3. According to the assumption $(A5)$, we have

$$\begin{aligned}\mathrm{E}_{t_1\cdots t_N}\Phi(Z_t) =&\mathrm{E}_{t_1\cdots t_N}\left(\sup_{f\in\mathcal{F}}\Big|\mathrm{E}_t f(Z_t) - \frac{1}{N}\sum_{n=1}^{N} f(Z_{t_n})\Big|\right) \\ =&\mathrm{E}_{t_1\cdots t_N}\left(\Big|\mathrm{E}_t f_t(Z_t) - \frac{1}{N}\sum_{n=1}^{N} f_t(Z_{t_n})\Big|\right) \\ \leq&\frac{1}{N}\mathrm{E}_{t_1\cdots t_N}\sum_{n=1}^{N}\mathrm{E}_t\Big|f_t(Z_t) - f_{t_n}(Z_{t_n})\Big| \\ \leq&\frac{\beta}{N}\mathrm{E}_{t_1\cdots t_N}\sum_{n=1}^{N}\mathrm{E}_t\Big\|Z_t - Z_{t_n}\Big\| \\ =&\frac{\beta}{N}\sum_{n=1}^{N}\mathrm{E}_{t,t_n}\Big\|Z_t - Z_{t_n}\Big\|.\end{aligned} \tag{48}$$

Since Lévy processes have stationary and independent increments, according to (48), we have that

$$\mathrm{E}_{t_1\cdots t_N}\Phi(Z_t) \leq \frac{\beta}{N}\sum_{n=1}^{N}\mathrm{E}_{|t-t_n|}\Big\|Z_{|t-t_n|}\Big\|. \tag{49}$$

This completes the proof. □

4.3 Proof of Theorem 3.5

Proof of Theorem 3.5.

According to (23), for any $c,t>0$, we have that

$$h_c(t) = \sqrt{\frac{tV(h_c(t))}{c}}. \tag{50}$$

According to (11) and (31), we have that

$$bh_c(t) = b\sqrt{\frac{tV(h_c(t))}{c}} \leq b\sqrt{\frac{tC_\nu}{c}}.$$

Since $b=o(\sqrt{c})$ as $N\to\infty$, we have that $bh_c(t)=o(1)$ as $N\to\infty$.

Similarly, according to (11), (31) and (50), we have that

$$\lambda cKh_c(t) = \lambda cK\sqrt{\frac{tV(h_c(t))}{c}} \leq \lambda K\sqrt{tC_\nu}\sqrt{c}. \tag{51}$$

Since λ and K are given constants, $\lambda cKh_c(t) = O(\sqrt{c})$ as $N\to\infty$, wherein $O(\cdot)$ stands for the infinitesimal of the same order, and thus we have $\lambda cKh_c(t)\leq o(1)$ $(N\to\infty)$ because $c=o(1)$. Finally, by using the condition that $\sum_{n=1}^{N} h_1(|t-t_n|) = o(N)$, we complete the proof. □

5 Conclusion

We present the risk bounds for Lévy processes without Gaussian components in the PAC-learning framework.

We first briefly introduce Lévy processes and discuss the relationship between infinitely divisible distributions and Lévy processes, which is necessary for the proof of the main result. Concentration inequalities are the main tools to achieve risk bounds for some empirical processes. Houdré and Marchal (2008) proposed a concentration inequality for Lévy processes without Gaussian components. However, this concentration inequality needs the condition that the function is 1-Lipschitz continuous. Thus, we first generalize it to the λ-Lipschitz ($\lambda > 0$) continuous version according to the discussion in (Houdré, 2002) and (Houdré and Marchal, 2008). Then, by using a concentration inequality for infinitely divisible distributions (Houdré, 2002), we prove that, with a high probability, the function $\Phi(Z_t)$ is Lipschitz continuous with respect to Z_t. Therefore, based on these results, we achieve the risk bounds for the Lévy processes without Gaussian components. By using the bound of fluctuations for Lévy processes (Houdré and Marchal, 2008), we use the resulted risk bound to obtain a convergence theorem which shows that the convergence of the risk bound is mainly determined by the Lévy measure ν. That implies that, if many large jumps appear frequently in the path of a Lévy process, the risk bound for the process may not converge to *zero* when the sample number approaches to *infinity*. In our future work, we will attempt to study risk bounds for other stochastic processes via concentration inequalities, *e.g.*, stochastic processes with exchangeable increments that are a well-known generalization of stochastic processes with independent increments (Kallenberg, 1973; Kallenberg, 1975).

Acknowledgements

The project was fully supported by the Nanyang Technological University SUG Grant (under project number M58020010).

References

L. G. Valiant (1984). *A theory of the learnable.* New York: ACM Press.

V. V. Petrov (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables.* USA: Oxford University Press.

V. N. Vapnik (1999). *The Nature of Statistical Learning Theory (Information Science and Statistics).* New York: Springer.

P. L. Bartlett, O. Bousquet and S. Mendelson (2002). Localized Rademacher Complexities, *In COLT 02: Proceedings of the 15th Annual Conference on Computational Learning Theory*, 44-58.

K. Sridharan, N. Srebro and S. Shalev-Shwartz (2008). Fast Rates for Regularized Objectives, *In NIPS 08: Neural Information Processing Systems.*

A. Tsybakov (2004). Optimal Aggregation of Classifiers in Statistical Learning. *Annals of Statistics* **32**:135-166.

O. Bousquet (2002). Concentration Inequalities and Empirical Processes Theory Applied to the Analysis of Learning Algorithms. *PhD thesis, Ecole Polytechnique.*

M. Mohri and A. Rostamizadeh (2008). Rademacher complexity bounds for non-i.i.d. processes. *In NIPS 08: Neural Information Processing Systems*, 1097-1104, Vancouver, Canada: MIT Press.

B. Yu (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Annals of Probability*, **22**(1):94-116.

A. B. Nobel and A. Dembo (1993). A note on uniform laws of averages for dependent processes. *Statistics and Probability Letters*, **17**:169-172.

O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick (2001). *Lévy Processes: Theory and Applications.* Birkhauser.

D. Applebaum (2004a). Lévy processes - from probability theory to finance and quantum groups. *Notices of the American Mathematical Society*, **51**(11):1336-1347.

D. Applebaum (2004b). *Lévy processes and and Stochastic Calculus.* Cambridge: Cambridge Press.

K. Sato (2004). *Lévy processes and infinitely divisible distributions.* USA: Cambridge University Press.

J. E. Figueroa-López and C. Houdré (2006). Risk bounds for the non-parametric estimation of Lévy processes. *IMS Lecture Notes-Monograph Series High Dimensional Probability*, **51**:96-116.

L. G. Valiant (1984). A Theory of the Learnable. *Commun. ACM*, **27**(11):1134-1142.

C. Houdré and P. Marchal (2008). Median, concentration and fluctuations for Lévy processes. *Stochastic Processes and their Applications*, **118**(5):852-863.

C. Houdré (2002). Remarks on deviation inequalities for functions of infinitely divisible random vectors. *Annals of Probability*, **30**(3):1223-1237.

M. B. Marcus and J. Rosinski (2001). L^1-Norm of Infinitely Divisible Random Vectors and Certain Stochastic Integrals. *Electronic Communications in Probability,* **6**:15-29.

O. Kallenberg (1973). Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrsch. Verw. Gebiete*, **27**:23-36.

O. Kallenberg (1975). On symmetrically distributed random measures. *Trans. Amer. Math. Soc.*, **202**:105-121.