Generalization Bound for Infinitely Divisible Empirical Process

Chao Zhang
School of Computer Engineering
Nanyang Technological University

Dacheng Tao
Centre for Quantum Computation
& Intelligent Systems (QCIS), FEIT
University of Technology, Sydney

Abstract

In this paper, we study the generalization bound for an empirical process of samples independently drawn from an infinitely divisible (ID) distribution, which is termed as the ID empirical process. In particular, based on a martingale method, we develop deviation inequalities for the sequence of random variables of an ID distribution. By applying the obtained deviation inequalities, we then show the generalization bound for ID empirical process based on the annealed Vapnik-Chervonenkis (VC) entropy. Afterward, according to Sauer's lemma, we get the generalization bound for ID empirical process based on the VC dimension. Finally, by using a resulted result bound, we analyze the asymptotic convergence of ID empirical process and show that the convergence rate of ID empirical process can reach $O\Big(\big(\frac{\Lambda_{\mathcal{F}}(2N)}{N}\big)^{\frac{1}{1.3}}\Big)$ and it is faster than the results of the generic i.i.d. empirical process (Vapnik, 1999).

1 Introduction

A probability distribution is said to be infinitely divisible if and only if it can be represented as the distribution of the sum of an arbitrary number of independently and identically distributed (i.i.d.) random variables. Infinitely divisible (ID) distribution covers lots of probability distributions including Poisson, geometric, lognormal, noncentral chi-square, exponential, Gamma, Pareto and Cauchy (Bose *et al.*, 2001). Therefore, ID distribution has a great theoretical value in probability and statistics.

Appearing in Proceedings of the 14^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2011, Fort Lauderdale, FL, USA. Volume 15 of JMLR: W&CP 15.

Moreover, many practical problems are related to ID distribution, *e.g.*, finance (Heston, 2004; Moosbrucker, 2007) and natural image statistics (Mumford and Gidas, 2001; Chainais, 2007). Some of these practical problems can be summed up as an empirical process of samples independently drawn from an ID distribution, *i.e.*, ID empirical process. Therefore, it is necessary to consider the asymptotical behavior of such empirical process, when the number of samples goes to the *infinity*. The generalization bound is the main method to study the asymptotical behavior of an empirical process (Vapnik, 1999; van der Vaart and Wellner, 1996).

Let $\mathcal{Z} := (\mathcal{X}, \mathcal{Y}) \subseteq \mathbb{R}^K$ be a space with $K = I + J$, where $\mathcal{X} \subseteq \mathbb{R}^I$ is the input space and $\mathcal{Y} \subseteq \mathbb{R}^J$ is the corresponding output space. It is expected to find a function $g^* : \mathcal{X} \to \mathcal{Y}$ such that for any $x \in \mathcal{X}$, $g^*(x)$ can precisely estimate the output $y \in \mathcal{Y}$. This can be achieved by minimizing the expected risk

$$\mathrm{E}(\ell(g(x), y)) := \int \ell(g(x), y) dP(\mathbf{z}), \tag{1}$$

where $\ell : \mathcal{Y}^2 \to \mathbb{R}$ is a loss function and $P(\mathbf{z})$ stands for the distribution of $\mathbf{z} = (x, y) \in \mathcal{Z}$. Since the distribution $P(\mathbf{z})$ is unknown, the target g^* usually cannot be directly obtained by minimizing (1). Instead, we introduce a function class $\mathcal{G}$ composed of real-valued functions defined on $\mathcal{Z}$ and a sample set $\mathbf{Z}_1^N := \{\mathbf{z}_n\}_{n=1}^N \subset \mathcal{Z}$ with $\mathbf{z}_n = (x_n, y_n)$. Given a function $g \in \mathcal{G}$, the empirical risk is defined as

$$\mathrm{E}_N(\ell(g(x), y)) := \frac{1}{N} \sum_{n=1}^{N} \ell(g(x_n), y_n), \tag{2}$$

which is regarded as an approximation of the expected risk (1). Alternatively, we minimize the empirical risk to obtain an estimate to g^*. We then define the loss function class

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

To simplify the presentation, for any $f \in \mathcal{F}$, we define

$$\mathrm{E}f := \int f(\mathbf{z}) dP(\mathbf{z}) \quad \text{and} \quad \mathrm{E}_N f := \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{z}_n).$$

The upper bound of $\sup_{f\in\mathcal{F}}\left|\mathrm{E}_N f-\mathrm{E}f\right|$ is the generalization bound for the empirical process indexed by the function class $\mathcal{F}$ and is one of major concerns in statistical learning theory. The generalization bound for an empirical process measures the probability that a function, chosen from a function class by an algorithm, has a sufficiently small error. In general, in order to obtain the generalization bound of a certain empirical process, one has to consider the following three key points: complexity measures of function classes, deviation (or concentration) inequalities and symmetrization inequalities for the empirical process.

There have been some generalization bounds obtained by using the concentration inequalities and the symmetrization inequalities for the generic i.i.d. empirical process. Vapnik (1999) gave generalization bounds based on the annealed VC entropy and the VC dimension, respectively. van der Vaart and Wellner (1996) showed the generalization bounds based on Rademacher complexities and covering numbers. Bartlett *et al.* (2005) developed the local Rademacher complexity and obtained a sharp error bound for a particular function class $\{f\in\mathcal{F}:\mathrm{E}f^2<\alpha\mathrm{E}f,\alpha>0\}$. Mohri and Rostamizadeh (2008) studied the generalization bound based on the Rademacher complexity for stationary β-mixing sequence. Zhang and Tao (2010) discussed generalization bounds for the Lévy process without Gaussian components.

In this paper, we study the generalization bounds for ID empirical process, where samples are independently drawn from an ID distribution. Although ID empirical process is a special case of the generic i.i.d. empirical process, it is still meaningful to study generalization bounds for ID empirical process and its signification can be summarized as follows:

- As mentioned above, since ID distribution covers a large body of probability distributions and many practical problems can be summed up as an empirical learning process based on ID distribution, it is necessary to investigate the asymptotic behavior of ID empirical process.
- In order to obtain the desired generalization bounds, new deviation inequalities have to be developed for ID empirical process and they are different from those for the generic i.i.d. empirical process.
- Because of the particularity of ID empirical process, the resulted generalization bounds have some specific properties that are different from the generalization bounds for the generic i.i.d. empirical process. We show that the convergence rate of ID empirical process can reach $O\Big(\big(\frac{\Lambda_{\mathcal{F}}(2N)}{N}\big)^{\frac{1}{1.3}}\Big)$ and it is faster than the results of the generic i.i.d. empirical process (Vapnik, 1999).

In order to obtain generalization bounds for ID empirical process, it is necessary to obtain suitable concentration (or deviation) inequalities. Houdré (2002) has proposed deviation inequalities for ID distribution. However, his results are only valid for one random variable of a special ID distribution, whose Gaussian component is *zero*. Therefore, his results cannot be directly used for the sequence of random variables of generic ID distribution.

Based on a martingale method, we extend Houdré's results (Houdré, 2002) and obtain deviation inequalities for the sequence of ID random variables. By using the resulted deviation inequalities, we then obtain the generalization bound for ID empirical process based on the annealed VC entropy. Afterward, according to Sauer's lemma (Sauer, 1972), we get the generalization bound based on the VC dimension. Finally, we analyze the asymptotic convergence of ID empirical process by using the resulted generalization bound.

The rest of this paper is organized as follows. Section 2 introduces ID distribution. Some deviation inequalities are presented in Section 3. We give generalization bounds for ID empirical process in Section 4. Some proofs of main results are shown in Section 5 and the last section concludes the paper.

2 Infinitely Divisible Distributions

In this section, we introduce some preliminaries on infinitely divisible (ID) distribution and please refer to (Sato, 2004) for details.

The ID distribution can be defined based on the characteristic function as follows:

Definition 2.1 *Let $\phi(t)$ be the characteristic function of a random variable $\mathbf{z}$*

$$\phi(t):=\mathrm{E}\left\{\mathrm{e}^{it\mathbf{z}}\right\}=\int_{-\infty}^{+\infty}\mathrm{e}^{it\mathbf{z}}dP(\mathbf{z}). \tag{3}$$

The distribution of $\mathbf{z}$ is infinitely divisible if and only if for any $N\in\mathbb{N}$, there exists a characteristic function $\phi_N(t)$ such that

$$\phi(t)=\underbrace{\phi_N(t)\times\cdots\times\phi_N(t)}_{N}, \tag{4}$$

where "$\times$" stands for multiplication.

According to the definition, if a random variable has the infinite divisibility, it can be represented as the sum of an arbitrary number of i.i.d. random variables.

Next, we introduce the Lévy measure and then show the characteristic exponent of an infinitely divisible distribution (Sato, 2004).

Definition 2.2 *Let ν be a Borel measure defined on $\mathbb{R}^K/\{0\}$. Then, the ν is said to be a Lévy measure if*

$$\int_{\mathbb{R}^K/\{0\}} \min\{\|\mathbf{u}\|^2, 1\}\nu(d\mathbf{u}) < \infty, \tag{5}$$

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

The Lévy measure describes the expected number of a certain height jump in a time interval of unit length. The characteristic exponent of an ID random variable is given by the following theorem (Sato, 2004).

Theorem 2.3 *(Lévy-Khintchine) A Borel probability measure μ of a random variable $\mathbf{z} \in \mathbb{R}^K$ is infinitely divisible if and only if there exists a triplet $(\mathbf{a}, \Sigma, \nu)$ such that for all $\theta \in \mathbb{R}^K$, the characteristic exponent $\ln \phi_\mu$ is of the form*

$$\begin{aligned}\ln\phi_\mu(\theta) =& i\langle \mathbf{a}, \theta\rangle - \frac{1}{2}\langle \theta, \Sigma\theta\rangle \\ &+ \int_{\mathbb{R}^K/\{0\}} \left(\mathrm{e}^{i\langle\theta,u\rangle} - 1 - i\langle\theta, \mathbf{u}\rangle \mathbf{1}_{\|\mathbf{u}\|\leq 1}\right)\nu(d\mathbf{u}),\end{aligned} \tag{6}$$

where $\mathbf{a} \in \mathbb{R}^K$, Σ is a $K \times K$ positive-definite symmetric matrix, ν is a Lévy measure on $\mathbb{R}^K/\{0\}$, and $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ stand for the inner product and a norm in $\mathbb{R}^K$, respectively.

Theorem 2.3 shows that an ID distribution can be completely determined by a triplet $(\mathbf{a}, \Sigma, \nu)$, where "$\mathbf{a}$" is the drift of a Brownian motion, "Σ" is a Gaussian component and "ν" is a Lévy measure. Thus, we call $(\mathbf{a}, \Sigma, \nu)$ the generating triplet of an ID distribution. The random variable of such ID distribution is denoted as the random variable with $(\mathbf{a}, \Sigma, \nu)$.

3 Deviation Inequality for Sequence of Infinitely Divisible Random Variables

Houdré (2002) gave deviation inequalities for one single ID random variable with the generating triplet $(\mathbf{a}, 0, \nu)$. However, his results are unsuitable to a sequence of ID random variables with $(\mathbf{a}, \Sigma, \nu)$.

In this section, we utilize a martingale method to extend Houdré's results and then develop the deviation inequalities for the sequence of ID random variables with $(\mathbf{a}, \Sigma, \nu)$, where Σ can be *nonzero*. Note that some mild conditions are required for the following discussion.

$(C1)$ The f is a partially differentiable function on $\mathbb{R}^K$ and there exists a constant $\beta_1 > 0$ such that for any $\mathbf{z} \in \mathbb{R}^K$,

$$\max_{1\leq k\leq K} \left|\frac{\partial f(\mathbf{z})}{\partial z_k}\right| \leq \beta_1.$$

$(C2)$ Denoting $\Sigma = \{\sigma_{ij}\}_{K\times K}$, there exists a constant $\beta_2 > 0$ such that

$$\max_{1\leq i,j\leq K} |\sigma_{ij}| \leq \beta_2. \tag{7}$$

$(C3)$ The ν has bounded support with $R = \inf\{\rho > 0 : \nu(\{\mathbf{u} : \|\mathbf{u}\| > \rho\}) = 0\}$.

The condition $(C1)$ implies that f has bounded partial derivatives. The constant β_1 is determined by the selected function and thus it is manipulatable. The condition $(C2)$ implies that all entries of Σ are bounded. The condition $(C3)$ implies the Lévy measure ν has a bounded support. We can use the nonparametric method proposed in (Watteel and Kulperger, 2003) to estimate β_2 and R. Next, we present a deviation inequality for a sequence of random variables with $(\mathbf{a}, \Sigma, \nu)$.

Theorem 3.1 *Assume that f is a function satisfying the condition $(C1)$. Let $\mathbf{Z}_1^N = \{\mathbf{z}_n\}_{n=1}^N$ be a sample set independently drawn from an ID distribution with $(\mathbf{a}, \Sigma, \nu)$ satisfying the condition $(C2)$. If $\mathrm{E}\mathrm{e}^{t\|\mathbf{z}\|} < +\infty$ holds for some $t > 0$, then we have for all $0 < \xi < \tau\big((M/\beta_1)^-\big)$,*

$$\Pr\left\{\left|F\left(\mathbf{Z}_1^N\right) - \mathrm{E}F\right| > \xi\right\} \leq \exp\left\{-\int_0^\xi \tau^{-1}(s)ds\right\}, \tag{8}$$

where

$$F\left(\mathbf{Z}_1^N\right) := \sum_{n=1}^N f(\mathbf{z}_n), \tag{9}$$

$\tau(a^-)$ is the left-hand limit of τ at a, $M = \sup\{t \geq 0 : \mathrm{E}\mathrm{e}^{t\|\mathbf{z}\|} < +\infty\}$ and τ^{-1} is the inverse of

$$\begin{aligned}\tau(t) =& \beta_1^2\beta_2 K^2 t \\ &+ N\int_{\mathbb{R}^K} \beta_1\|\mathbf{u}\|\left(\mathrm{e}^{t\beta_1\|\mathbf{u}\|} - 1\right)\nu(d\mathbf{u}),\end{aligned} \tag{10}$$

with the domain of $0 < t < M/\beta_1$.

Because of an integral of τ^{-1}, the deviation inequality (8) cannot explicitly reflect the asymptotic behavior of $\Pr\left\{\left|F\left(\mathbf{Z}_1^N\right) - \mathrm{E}F\right| > \xi\right\}$ when N goes to the *infinity*. Thus, we introduce an extra condition that the Lévy measure ν has a bounded support, and then we obtain another deviation inequality for a sequence of random variables with $(\mathbf{a}, \Sigma, \nu)$.

Corollary 3.2 *Following notations in Theorem 3.1, let* $V = \int_{\mathbb{R}^K} \|\mathbf{u}\|^2 \nu(d\mathbf{u})$ *and* ν *satisfy the condition* $(C3)$. *Then, we have for any* $\xi > 0$,

$$\Pr\left\{\left|F\left(\mathbf{Z}_1^N\right) - \mathrm{E}F\right| > \xi\right\}$$
$$\leq \exp\left\{\frac{\xi}{\beta_1 R} - \left(\frac{\xi}{\beta_1 R} + \frac{N(\beta_1^2\beta_2 K^2 + V)}{\beta_1^2 R^2}\right) \times \ln\left(1 + \frac{\xi\beta_1 R}{N(\beta_1^2\beta_2 K^2 + V)}\right)\right\}. \quad (11)$$

Based on the above two deviation inequalities (8) and (11), we can obtain generalization bounds for ID empirical processes.

4 Generalization Bound for Infinitely Divisible Empirical Process

In this section, we utilize the resulted deviation inequalities to obtain the generalization bounds for ID empirical process based on the annealed VC entropy. Then, by using Sauer's lemma (Sauer, 1972), we obtain the generalization bounds based on the VC dimension. Since we are mainly concerned with the function class composed of real-valued functions satisfying the condition $(C1)$, the annealed VC entropy and the VC dimension are defined in the scenario of real function classes (Vapnik, 1999).

4.1 Complexity Measure for Function Class

Following the style of Vapnik's work (Vapnik, 1999), we can define the annealed VC entropy and the VC dimension as follows.

Definition 4.1 *Assume that* $\mathcal{F}$ *is a real function class composed of functions with the range* $[A, B]$ *and* $\mathbf{Z}_1^N = \{\mathbf{z}_n\}_{n=1}^N$ *is a sample set drawn from* $\mathcal{Z}$. *For any* $\alpha \in (A, B)$, *define*

$$\vec{f}(\alpha, \mathbf{Z}_1^N) := \big(f(\mathbf{z}_1) - \alpha, \cdots, f(\mathbf{z}_N) - \alpha\big), \quad (12)$$

and

$$\mathrm{Ind}\left(\vec{f}(\alpha, \mathbf{Z}_1^N)\right) := \big(\delta(f(\mathbf{z}_1) - \alpha), \cdots, \delta(f(\mathbf{z}_N) - \alpha)\big), \quad (13)$$

where δ *is an indicator function*

$$\delta(x) := \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

According to (12) *and* (13), *we obtain a set associated with* $\mathbf{Z}_1^N$

$$\mathcal{F}_{\mathbf{Z}_1^N} := \left\{\mathrm{Ind}\left(\vec{f}(\alpha, \mathbf{Z}_1^N)\right) : f \in \mathcal{F},\ \alpha \in (A, B)\right\}.$$

Then, the corresponding annealed VC entropy is defined as

$$\Lambda_{\mathcal{F}}(N) := \ln \mathrm{E}\left\{\left|\mathcal{F}_{\mathbf{Z}_1^N}\right|\right\},$$

where $\left|\mathcal{F}_{\mathbf{Z}_1^N}\right|$ *stands for the cardinality of* $\mathcal{F}_{\mathbf{Z}_1^N}$. *The VC dimension of* $\mathcal{F}$ *is defined as*

$$VC(\mathcal{F}) := \max\left\{N > 0 : \max_{\mathbf{Z}_1^N \in \mathcal{Z}^N} \left|\mathcal{F}_{\mathbf{Z}_1^N}\right| = 2^N\right\}.$$

Moreover, Sauer's lemma (Sauer, 1972) shows a relationship between the annealed VC entropy and the VC dimension.

Lemma 4.2 *(Sauer) Following notations in Definition 4.1, if* $VC(\mathcal{F}) \leq D$, *then we have*

$$\exp\{\Lambda_{\mathcal{F}}(N)\} \leq \sum_{d=0}^{D} \binom{N}{d}. \quad (14)$$

Furthermore, for any $N \geq D$, *we have*

$$\exp\{\Lambda_{\mathcal{F}}(N)\} \leq \left(\frac{\mathrm{e}N}{D}\right)^D. \quad (15)$$

Next, we present the main results of this paper.

4.2 Generalization Bound for Infinitely Divisible Empirical Process

In Theorem 3.1 and Corollary 3.2, we give the deviation inequalities for a sequence of ID random variables with $(\mathbf{a}, \Sigma, \nu)$. In order to achieve generalization bounds for ID empirical process, we need the following lemma and its proof is given in the next section.

Lemma 4.3 *Let* $\mathcal{F}$ *be a function class with the range* $[A, B]$ *and* $\mathbf{Z}_1^{2N} = \{\mathbf{z}_n\}_{n=1}^{2N}$ *be an i.i.d. sample set. Then, for any* $\xi > 0$ *such that* $N\xi^2 \geq 32\max\{A^2, B^2\}$, *we have*

$$\Pr\left\{\sup_{f\in\mathcal{F}} \left|\mathrm{E}f - \mathrm{E}_N f\right| > \xi\right\}$$
$$\leq 2\mathrm{E}\left\{\left|\mathcal{F}_{\mathbf{Z}_1^{2N}}\right|\right\} \max_{f\in\mathcal{F}} \Pr\left\{\left|\mathrm{E}f - \mathrm{E}_N f\right| > \frac{\xi}{4}\right\}. \quad (16)$$

This result provides a bridge between deviation inequalities and generalization bounds for an empirical process. Subsequently, we show generalization bounds for ID empirical process based on the annealed VC entropy and the VC dimension, respectively.

Theorem 4.4 *Assume that* $\mathcal{F}$ *is a function class composed of functions with the range* $[A, B]$ *and satisfying the condition* $(C1)$. *Let* $\mathbf{Z}_1^N = \{\mathbf{z}_n\}_{n=1}^N$ *be a* *sample set independently drawn from an ID distribution with* $(\mathbf{a}, \Sigma, \nu)$ *satisfying the condition* $(C2)$. *If* $\mathrm{E}e^{t\|\mathbf{z}\|} < +\infty$ *holds for some* $t > 0$*, then for all* $\xi > 0$ *such that* $0 < N\xi/4 < \tau\big((M/\beta_1)^-\big)$ *and* $N\xi^2 \geq 32\max\{A^2, B^2\}$*, we have*

$$\begin{aligned}&\Pr\left\{\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f - \mathrm{E}f\big| > \xi\right\}\\ \leq &2\exp\left\{\Lambda_{\mathcal{F}}(2N) - \int_0^{\frac{N\xi}{4}}\tau^{-1}(s)ds\right\}, \end{aligned} \tag{17}$$

where $\tau(a^-)$ *denotes the left-hand limit of* τ *at the point* a*,* $M = \sup\big\{t \geq 0 : \mathrm{E}e^{t\|\mathbf{z}\|} < +\infty\big\}$ *and* τ^{-1} *is the inverse of*

$$\tau(t) = \beta_1^2\beta_2K^2t + N\int_{\mathbb{R}^K}\beta_1\|\mathbf{u}\|\left(e^{t\beta_1\|\mathbf{u}\|} - 1\right)\nu(d\mathbf{u}),$$

with the domain of $0 < t < M/\beta_1$*. Furthermore, if* $VC(\mathcal{F}) \leq D$*, then there holds that for any* $N > D/2$*,*

$$\begin{aligned}&\Pr\left\{\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f - \mathrm{E}f\big| > \xi\right\}\\ \leq &2\exp\left\{D\ln\left(\frac{2eN}{D}\right) - \int_0^{\frac{N\xi}{4}}\tau^{-1}(s)ds\right\}. \end{aligned} \tag{18}$$

Proof. According to Theorem 3.1, Lemma 4.2 and Lemma 4.3, we can directly obtain (17) and (18). This completes the proof. ■

Since (17) and (18) are given by incorporating the integrals of τ^{-1}, the asymptotic behavior of the generalization bounds cannot be explicitly reflected, when N goes to the *infinity*. Moreover, the applicability of Theorem 4.4 is restricted by two conditions $0 < N\xi/4 < \tau\big((M/\beta_1)^-\big)$ and $N\xi^2 \geq 32\max\{A^2, B^2\}$. To overcome these drawbacks, we develop other generalization bounds for ID empirical process by adding a mild condition that the Lévy measure ν has a bounded support.

Theorem 4.5 *Following notations in Theorem 4.4, let* $V = \int_{\mathbb{R}^K}\|\mathbf{u}\|^2\nu(d\mathbf{u})$ *and* ν *satisfy the condition* $(C3)$*. Then, we have for any* $\xi > 0$ *such that* $N\xi^2 \geq 32\max\{A^2, B^2\}$*,*

$$\begin{aligned}&\Pr\left\{\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f - \mathrm{E}f\big| > \xi\right\}\\ \leq &2\exp\Big\{\Lambda_{\mathcal{F}}(2N) + \frac{N(\beta_1^2\beta_2K^2 + V)}{\beta_1^2R^2}\\ &\qquad\times\Gamma\left(\frac{\xi\beta_1R}{4(\beta_1^2\beta_2K^2 + V)}\right)\Big\}, \end{aligned} \tag{19}$$

where

$$\Gamma(x) = x - (x+1)\ln(x+1). \tag{20}$$

Furthermore, if $VC(\mathcal{F}) \leq D$*, then we have for any* $N > D/2$*,*

$$\begin{aligned}&\Pr\left\{\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f - \mathrm{E}f\big| > \xi\right\}\\ \leq &2\exp\Big\{D\ln\left(\frac{2eN}{D}\right) + \frac{N(\beta_1^2\beta_2K^2 + V)}{\beta_1^2R^2}\\ &\qquad\times\Gamma\left(\frac{\xi\beta_1R}{4(\beta_1^2\beta_2K^2 + V)}\right)\Big\}. \end{aligned} \tag{21}$$

Proof. This theorem can be directly obtained from Corollary 3.2, Lemma 4.2 and Lemma 4.3. This completes the proof. ■

Theorem 4.5 shows that if the Lévy measure ν has the bounded support, we can obtain the generalization bounds that can explicitly reflect the asymptotic behavior of ID empirical process, when N goes to the *infinity*. By combining (20) and (21), we can obtain the following theorem.

Theorem 4.6 *Assume that the conditions* $(C1)$*-*$(C3)$ *are all valid. Let* x^* *be the solution of the equation*

$$\Gamma(x) = x - (x+1)\ln(x+1) = 0.$$

If $VC(\mathcal{F}) \leq D$ *and*

$$\lim_{N\to\infty}\frac{\ln(N/D)}{(N/D)} = 0,$$

then for any $\xi > \frac{4x^*(\beta_1^2\beta_2K^2+V)}{\beta_1R}$*, we have*

$$\lim_{N\to\infty}\Pr\left\{\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f - \mathrm{E}f\big| > \xi\right\} = 0.$$

As shown in this theorem, if the VC dimension for $\mathcal{F}$ is finite, there holds that for some $\xi > 0$, $\Pr\big\{\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f - \mathrm{E}f\big| > \xi\big\}$ converges to *zero*, when the number of samples goes to the *infinity*. This is partly in accordance with Vapnik's results (Vapnik, 1999). However, as shown in (19) and (21), because of the particularity of ID empirical process, the convergence rate of $\sup_{f\in\mathcal{F}}\big|\mathrm{E}_N f - \mathrm{E}f\big|$ is faster than the case of the generic i.i.d. empirical process. The detailed discussion on the convergence rate is postponed in the appendix.

5 Proofs of Main Results

In this section, we prove Theorem 3.1, Corollary 3.2 and Lemma 4.3, respectively.

5.1 Martingale Method

In this paper, we use the following martingale method to extend Houdré's deviation inequalities (Houdré, 2002) to the sequence of ID random variables.

For any $0 \leq m \leq N$, define a random variable

$$S_m := \mathrm{E}\left\{F(\mathbf{Z}_{n=1}^N)|\mathbf{Z}_1^m\right\}, \tag{22}$$

where $\mathbf{Z}_1^m = \{\mathbf{z}_1, \cdots, \mathbf{z}_m\} \subseteq \mathbf{Z}_1^N$ and $\mathbf{Z}_1^0 = \varnothing$. It is direct that $S_0 = \mathrm{E}F$ and $S_N = F(\mathbf{Z}_1^N)$.

According to (9) and (22), for any $1 \leq m \leq N$, letting

$$\psi_m(\mathbf{Z}_1^N) := S_m - S_{m-1}, \tag{23}$$

we have

$$\begin{aligned}
\psi_m(\mathbf{Z}_1^N) =& \mathrm{E}\left\{F(\mathbf{Z}_1^N)|\mathbf{Z}_1^m\right\} - \mathrm{E}\left\{F(\mathbf{Z}_1^N)|\mathbf{Z}_1^{m-1}\right\} \\
=& \mathrm{E}\left\{\sum_{n=1}^N f(\mathbf{z}_n)\Big|\mathbf{Z}_1^m\right\} \\
&- \mathrm{E}\left\{\sum_{n=1}^N f(\mathbf{z}_n)\Big|\mathbf{Z}_1^{m-1}\right\} \\
=& \sum_{n=1}^m f(\mathbf{z}_n) + \mathrm{E}\left\{\sum_{n=m+1}^N f(\mathbf{z}_n)\right\} \\
&- \left(\sum_{n=1}^{m-1} f(\mathbf{z}_n) + \mathrm{E}\left\{\sum_{n=m}^N f(\mathbf{z}_n)\right\}\right) \\
=& f(\mathbf{z}_m) - \mathrm{E}f(\mathbf{z}_m),
\end{aligned} \tag{24}$$

and thus

$$\mathrm{E}\left\{\psi_m(\mathbf{Z}_1^N)\big|\mathbf{Z}_1^{m-1}\right\} = \mathrm{E}\left\{\psi_m(\mathbf{Z}_1^N)\right\} = 0. \tag{25}$$

Moreover, we also have the following lemma.

Lemma 5.1 *Following the notation in* (9) *and* (23)*, we have*

$$\Pr\left\{F\left(\mathbf{Z}_1^N\right) - \mathrm{E}F > \xi\right\} \leq \mathrm{e}^{-t\xi} \prod_{m=1}^N \mathrm{E}\left\{\mathrm{e}^{t\psi_m}\right\}. \tag{26}$$

Proof. According to (24), Markov's inequality and the law of iterated expectation, we have

$$\begin{aligned}
&\Pr\left\{F\left(\mathbf{Z}_1^N\right) - \mathrm{E}F > \xi\right\} \\
\leq& \mathrm{e}^{-t\xi}\mathrm{E}\left\{\mathrm{e}^{t\left(F(\mathbf{Z}_1^N) - \mathrm{E}F\right)}\right\} \\
=& \mathrm{e}^{-t\xi}\mathrm{E}\left\{\mathrm{e}^{t\sum_{n=1}^N (S_m - S_{m-1})}\right\} \\
=& \mathrm{e}^{-t\xi}\mathrm{E}\left\{\mathrm{E}\left\{\mathrm{e}^{t\sum_{m=1}^N (S_m - S_{m-1})}\Big|\mathbf{Z}_1^{N-1}\right\}\right\} \\
=& \mathrm{e}^{-t\xi}\mathrm{E}\left\{\mathrm{e}^{t\sum_{m=1}^{N-1} (S_m - S_{m-1})}\mathrm{E}\left\{\mathrm{e}^{t(S_N - S_{N-1})}\Big|\mathbf{Z}_1^{N-1}\right\}\right\} \\
=& \mathrm{e}^{-t\xi}\prod_{n=1}^N \mathrm{E}\left\{\mathrm{e}^{t(S_m - S_{m-1})}\Big|\mathbf{Z}_1^{m-1}\right\} \\
=& \mathrm{e}^{-t\xi}\prod_{m=1}^N \mathrm{E}\left\{\mathrm{e}^{t\psi_m}\right\}.
\end{aligned} \tag{27}$$

This completes the proof. ∎

5.2 Proofs of Theorem 3.1 and Corollary 3.2

First, we give some preliminaries on the functions ψ_m ($1 \leq m \leq N$). Let $\mathbf{W}_1^N = \{\mathbf{w}_n\}_{n=1}^N \subset \mathbb{R}^K$, and then according to (23), we have for any $1 \leq m \leq N$,

$$\begin{aligned}
\nabla\psi_m(\mathbf{W}_1^N) &= \nabla(f(\mathbf{w}_m) - \mathrm{E}f(\mathbf{w}_m)) \\
&= \nabla f(\mathbf{w}_m),
\end{aligned} \tag{28}$$

and

$$\begin{aligned}
\nabla\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)} =& t\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\nabla\psi_m(\mathbf{W}_1^N) \\
=& t\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\nabla f(\mathbf{w}_m),
\end{aligned} \tag{29}$$

where ∇ is the gradient operator.

Let $\mathbf{u}_1^N := \{\mathbf{u}_m\}_{m=1}^N \subset \mathbb{R}^K$. According to the Mean-Value Theorem for multivariate functions (Courant and John, 1974), if f satisfies the condition $(C1)$, we have

$$\begin{aligned}
&\psi_m(\mathbf{W}_1^N + \mathbf{u}_1^N) - \psi_m(\mathbf{W}_1^N) \\
=& f(\mathbf{w}_m + \mathbf{u}_m) - f(\mathbf{w}_m) \\
=& \langle\nabla f(\widetilde{\mathbf{w}}_m), \mathbf{u}_m\rangle \leq K\beta_1\|\mathbf{u}_m\| \\
\leq& K\beta_1\|\mathbf{u}_1^N\|,
\end{aligned} \tag{30}$$

where $\widetilde{\mathbf{w}}_m$ is an intermediate point on the line segment between the two points $\mathbf{w}_m + \mathbf{u}_m$ and $\mathbf{w}_m$.

We also need the following result given in (Houdré *et al.*, 1998).

Lemma 5.2 *Let* $\mathbf{z}$ *be drawn from an ID distribution with the generating triplet* $(\mathbf{a}, \Sigma, \nu)$ *such that* $\mathrm{E}\|\mathbf{z}\|^2 < +\infty$*. If* $f, g : \mathbb{R}^K \to \mathbb{R}$ *are partially differentiable functions, then*

$$\mathrm{E}f(\mathbf{z})g(\mathbf{z}) - \mathrm{E}f(\mathbf{z})\mathrm{E}g(\mathbf{z}) = \int_0^1 \mathrm{E}_z\Big\{\langle\Sigma\nabla f(\mathbf{z}), \nabla g(\mathbf{w})\rangle$$
$$+ \int_{\mathbb{R}^K}\big(f(\mathbf{z}+\mathbf{u}) - f(\mathbf{z})\big)\big(g(\mathbf{w}+\mathbf{u}) - g(\mathbf{w})\big)\nu(d\mathbf{u})\Big\}dz,$$

where the expectation E_z *is signified in Proposition 2 in (Houdré et al., 1998).*

In the proofs of Theorem 3.1 and Corollary 3.2, we adopt some techniques appearing in Houdré's work (Houdré, 2002).

Proof of Theorem 3.1. First, we consider the validity of our proof. According to Theorem 25.3 in (Sato, 2004), we have

$$\Omega = \Big\{\alpha \geq 0 : \mathrm{E}\mathrm{e}^{\alpha\|\mathbf{z}\|} < +\infty\Big\}$$

$$=\left\{\alpha \geq 0 : \int_{\|\mathbf{u}\|>1} \mathrm{e}^{\alpha\|\mathbf{u}\|}\nu(d\mathbf{u}) < +\infty\right\}$$
$$=\left\{\alpha \geq 0 : \int_{\|\mathbf{u}\|>1} \left(\mathrm{e}^{\alpha\|\mathbf{u}\|} - \alpha\|\mathbf{u}\| - 1\right)\nu(d\mathbf{u}) < +\infty\right\},$$

which implies that Ω is an interval and not reduced to $\{0\}$. Thus, the following discussion is valid.

Define $\mathbf{u}_1^N := \{\mathbf{u}_n\}_{n=1}^N \subset \mathbb{R}^K$ and let $\mathbf{W}_1^N = \{\mathbf{w}_n\}_{n=1}^N$ be another sample set independently drawn from the ID distribution $\mathcal{Z}$. By combining (24), (25), (28), (29), (30) and Lemma 5.2, for any $1 \leq m \leq N$, we have

$$\begin{aligned}
&\mathrm{E}\left\{\psi_m(\mathbf{Z}_1^N)\mathrm{e}^{t\psi_m(\mathbf{Z}_1^N)}\right\} - \mathrm{E}\left\{\psi_m(\mathbf{Z}_1^N)\right\}\mathrm{E}\left\{\mathrm{e}^{t\psi_m(\mathbf{Z}_1^N)}\right\}\\
=&\int_0^1 \mathrm{E}_z\Big\{\left\langle \Sigma\nabla\psi_m(\mathbf{Z}_1^N), \nabla \mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\right\rangle\\
&+\int \Big(\psi_m(\mathbf{Z}_1^N+\mathbf{u}_1^N)-\psi_m(\mathbf{Z}_1^N)\Big)\\
&\times\Big(\mathrm{e}^{t\psi_m(\mathbf{W}_1^N+\mathbf{u}_1^N)} - \mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\Big)\nu(d\mathbf{u}_1^N)\Big\}dz\\
=&\int_0^1 \mathrm{E}_z\Big\{\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\left\langle \Sigma\nabla f(\mathbf{z}_m), \nabla f(\mathbf{w}_m)\right\rangle\\
&+\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\int \Big(\psi_m(\mathbf{Z}_1^N+\mathbf{u}_1^N)-\psi_m(\mathbf{Z}_1^N)\Big)\\
&\times\Big(\mathrm{e}^{t\left(\psi_m(\mathbf{W}_1^N+\mathbf{u}_1^N)-\psi_m(\mathbf{W}_1^N)\right)} - 1\Big)\nu(d\mathbf{u}_1^N)\Big\}dz\\
\leq&\int_0^1 \mathrm{E}_z\Big\{\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\Big\}\Big(\beta_1^2\beta_2K^2t\\
&+\int_{\mathbb{R}^K}\beta_1\|\mathbf{u}_m\|\left(\mathrm{e}^{t\beta_1\|\mathbf{u}_m\|}-1\right)\nu(d\mathbf{u}_m)\Big)dz\\
=&\mathrm{E}\Big\{\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}\Big\}\Big(\beta_1^2\beta_2K^2t\\
&+\int_{\mathbb{R}^K}\beta_1\|\mathbf{u}_m\|\left(\mathrm{e}^{t\beta_1\|\mathbf{u}_m\|}-1\right)\nu(d\mathbf{u}_m)\Big).
\end{aligned}\tag{31}$$

Since the marginal distribution of $(\mathbf{Z}_1^N, \mathbf{W}_1^N)$ is $\mathbf{Z}_1^N$ and $\mathbf{Z}_1^N$ has the same distribution as that of $\mathbf{W}_1^N$, letting $L(t) = \mathrm{E}\mathrm{e}^{t\psi_m(\mathbf{W}_1^N)}$, we have

$$\begin{aligned}
\frac{L'(t)}{L(t)} =& \frac{\mathrm{E}\psi_m\mathrm{e}^{t\psi_m(\mathbf{Z}_1^N)}}{\mathrm{E}\mathrm{e}^{t\psi_m(\mathbf{Z}_1^N)}}\\
\leq&\beta_1^2\beta_2K^2t + \int_{\mathbb{R}^K}\beta_1\|\mathbf{u}_m\|\left(\mathrm{e}^{t\beta_1\|\mathbf{u}_m\|}-1\right)\nu(d\mathbf{u}_m).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\int_0^t \frac{L'(s)}{L(s)}ds \leq& \int_0^t \Big(\beta_1^2\beta_2K^2s\\
&+\int_{\mathbb{R}^K}\beta_1\|\mathbf{u}_m\|\left(\mathrm{e}^{s\beta_1\|\mathbf{u}_m\|}-1\right)\nu(d\mathbf{u}_m)\Big)ds,
\end{aligned}\tag{32}$$

and then by (25),

$$\begin{aligned}
\ln \mathrm{E}\mathrm{e}^{s\psi_m}\Big|_0^t =& \ln \mathrm{E}\mathrm{e}^{t\psi_m} \leq \frac{\beta_1^2\beta_2K^2t^2}{2}\\
&+\int_{\mathbb{R}^K}\left(\mathrm{e}^{t\beta_1\|\mathbf{u}_m\|}-t\beta_1\|\mathbf{u}_m\|-1\right)\nu(d\mathbf{u}_m).
\end{aligned}\tag{33}$$

By combining (33) and Lemma 5.1, we have

$$\Pr\left\{F\left(\mathbf{Z}_1^N\right) - \mathrm{E}F > \xi\right\} \leq \mathrm{e}^{\Phi(t)-t\xi},\tag{34}$$

where

$$\begin{aligned}
\Phi(t) =& \frac{N\beta_1^2\beta_2K^2t^2}{2}\\
&+\sum_{m=1}^N\int_{\mathbb{R}^K}\left(\mathrm{e}^{t\beta_1\|\mathbf{u}_m\|}-t\beta_1\|\mathbf{u}_m\|-1\right)\nu(d\mathbf{u}_m)\\
=& \frac{N\beta_1^2\beta_2K^2t^2}{2}\\
&+N\int_{\mathbb{R}^K}\left(\mathrm{e}^{t\beta_1\|\mathbf{u}\|}-t\beta_1\|\mathbf{u}\|-1\right)\nu(d\mathbf{u}).
\end{aligned}\tag{35}$$

Since $\mathrm{E}\mathrm{e}^{t\|\mathbf{z}\|} < +\infty$, for all $0 < t < M$, Φ is infinitely differentiable on $(0, M)$ with

$$\begin{aligned}
\Phi'(t) = \tau(t) =& N\beta_1^2\beta_2K^2t\\
&+N\int_{\mathbb{R}^K}\beta_1\|\mathbf{u}\|\left(\mathrm{e}^{t\beta_1\|\mathbf{u}\|}-1\right)\nu(d\mathbf{u}) > 0,
\end{aligned}\tag{36}$$

and

$$\begin{aligned}
\Phi''(t) =& N\beta_1^2\beta_2K^2\\
&+N\int_{\mathbb{R}^K}\beta_1^2\|\mathbf{u}\|^2\mathrm{e}^{t\beta_1\|\mathbf{u}\|}\nu(d\mathbf{u}) > 0.
\end{aligned}\tag{37}$$

Then, we minimize the right-hand side of (34) with respect to t. According to (36) and (37), for any $0 < \xi < \tau(M^{-1})$, $\min_{0<t<M}\{\Phi(t) - t\xi\}$ is achieved when $\tau(t) - \xi = 0$. Since $\Phi(0) = \tau(0) = \tau^{-1}(0) = 0$, we have

$$\begin{aligned}
\Phi\left(\tau^{-1}(\xi)\right) =& \int_0^{\tau^{-1}(\xi)}\tau(s)ds = \int_0^\xi s d\tau^{-1}(s)\\
=&\xi\tau^{-1}(\xi) - \int_0^\xi \tau^{-1}(s)ds.
\end{aligned}\tag{38}$$

Thus, for any $0 < \xi < \tau(M^{-1})$,

$$\min_{0<t<M}\{\Phi(t) - t\xi\} = -\int_0^\xi \tau^{-1}(s)ds.$$

Similarly, we also can prove that

$$\Pr\left\{\mathrm{E}F - F\left(\mathbf{Z}_1^N\right) > \xi\right\} \leq \exp\left(-\int_0^\xi \tau^{-1}(s)ds\right).$$

This completes the proof. ■

Proof of Corollary 3.2. Since ν satisfies the condition $(C3)$, the support $supp(\nu) \subseteq [-R, R]$ and then $\mathrm{E}e^{t\|\mathbf{z}\|} < +\infty$ holds for any $t > 0$. Then, we have

$$\begin{aligned}\tau(t) =& N\beta_1^2\beta_2K^2t \\ &+ N\int_{\|\mathbf{u}\|\leq R}\beta_1\|\mathbf{u}\|\left(e^{t\beta_1\|\mathbf{u}\|}-1\right)\nu(d\mathbf{u}) \\ =& N\beta_1^2\beta_2K^2t + N\int_{\|\mathbf{u}\|\leq R}\beta_1^2\|\mathbf{u}\|^2 \\ &\times\left(\sum_{k=1}^{\infty}\frac{t^k\beta_1^{k-1}\|\mathbf{u}\|^{k-1}}{k!}\right)\nu(d\mathbf{u}) \\ \leq& N\beta_1^2\beta_2K^2t + N\int_{\|\mathbf{u}\|\leq R}\beta_1^2\|\mathbf{u}\|^2 \\ &\times\left(\sum_{k=1}^{\infty}\frac{t^k(\beta_1R)^{k-1}}{k!}\right)\nu(d\mathbf{u}) \\ =& N\beta_1^2\beta_2K^2t + NV\left(\frac{e^{t\beta_1R}-1}{\beta_1R}\right) \\ \leq& N(\beta_1^2\beta_2K^2+V)\left(\frac{e^{t\beta_1R}-1}{\beta_1R}\right). \end{aligned} \tag{39}$$

As shown in (36) and (37), $\tau(t)$ is an increasing function and thus $\tau^{-1}(t)$ is also an increasing function. Moreover, according to Theorem 3.1 and (39), we have for any $\xi > 0$,

$$\begin{aligned}&\Pr\left\{F(\mathbf{Z}_1^N)-\mathrm{E}F>\xi\right\} \\ \leq& \exp\left\{-\int_0^{\xi}\frac{1}{\beta_1R}\ln\left(1+\frac{\beta_1Rs}{N(\beta_1^2\beta_2K^2+V)}\right)ds\right\} \\ =& \exp\left\{\frac{\xi}{\beta_1R}-\left(\frac{\xi}{\beta_1R}+\frac{N(\beta_1^2\beta_2K^2+V)}{\beta_1^2R^2}\right)\right. \\ &\left.\times\ln\left(1+\frac{\xi\beta_1R}{N(\beta_1^2\beta_2K^2+V)}\right)\right\}. \end{aligned} \tag{40}$$

This completes the proof. ■

5.3 Proof of Lemma 4.3

Before the formal proof, we introduce the symmetrization inequality and its details are given in (Bousquet *et al.*, 2004).

Lemma 5.3 *(Symmetrization) Assume that $\mathcal{F}$ is a function class and let $\mathbf{Z}_1^N, \mathbf{Z}'^N_1$ be two i.i.d. sample sets. Then, for any $\xi > 0$ such that $N\xi^2 \geq 32\max\{A^2, B^2\}$, we have*

$$\begin{aligned}&\Pr\left\{\sup_{f\in\mathcal{F}}\left|\mathrm{E}f-\mathrm{E}_Nf\right|>\xi\right\} \\ \leq& 2\Pr\left\{\sup_{f\in\mathcal{F}}\left|\mathrm{E}'_Nf-\mathrm{E}_Nf\right|>\frac{\xi}{2}\right\}. \end{aligned} \tag{41}$$

Based on Lemma 5.3, we can prove Lemma 4.3.

Proof of Lemma 4.3. According to Lemma 5.3, for any $\xi > 0$ such that $N\xi^2 \geq 32\max\{A^2, B^2\}$, we have

$$\begin{aligned}&\Pr\left\{\sup_{f\in\mathcal{F}}\left|\mathrm{E}f-\mathrm{E}_Nf\right|>\xi\right\} \\ \leq& 2\Pr\left\{\sup_{f\in\mathcal{F}}\left|\mathrm{E}'_Nf-\mathrm{E}_Nf\right|>\frac{\xi}{2}\right\} \\ \leq& 2\mathrm{E}\left\{\left|\mathcal{F}_{\mathbf{Z}_1^{2N}}\right|\right\}\max_{f\in\mathcal{F}}\Pr\left\{\left|\mathrm{E}'_Nf-\mathrm{E}_Nf\right|>\frac{\xi}{2}\right\} \\ \leq& 2\mathrm{E}\left\{\left|\mathcal{F}_{\mathbf{Z}_1^{2N}}\right|\right\}\max_{f\in\mathcal{F}}\Pr\Big\{|\mathrm{E}f-\mathrm{E}'_Nf| \\ &+|\mathrm{E}f-\mathrm{E}_Nf|>\frac{\xi}{2}\Big\}. \end{aligned} \tag{42}$$

Since $\mathbf{Z}_1^N$ and $\mathbf{Z}'^N_1$ are both independently drawn from an identical distribution, according to (42), we have

$$\begin{aligned}&\Pr\left\{\sup_{f\in\mathcal{F}}\left|\mathrm{E}f-\mathrm{E}_Nf\right|>\xi\right\} \\ \leq& 2\mathrm{E}\left\{\left|\mathcal{F}_{\mathbf{Z}_1^{2N}}\right|\right\}\max_{f\in\mathcal{F}}\Pr\left\{\left|\mathrm{E}f-\mathrm{E}_Nf\right|>\frac{\xi}{4}\right\}. \end{aligned} \tag{43}$$

This completes the proof. ■

6 Conclusion

In this paper, we study the generalization bounds for the empirical process of samples independently drawn from an infinitely divisible (ID) distribution with the generating triplet $(\mathbf{a}, \Sigma, \nu)$. By using a martingale method, we provide two kinds of deviation inequalities for a sequence of random variables with $(\mathbf{a}, \Sigma, \nu)$. We then utilize the resulted deviation inequalities to obtain the generalization bounds based on the annealed VC entropy for ID empirical process. According to Sauer's lemma, we further obtain the generalization bounds based on the VC dimension, respectively. We find that the asymptotic convergence of the generalization bounds is determined by the complexity of the function class $\mathcal{F}$ measured by the annealed VC entropy or the VC dimension. This is in accordance with Vapnik's results on the asymptotic convergence for i.i.d. empirical processes. However, because of the particularity of ID empirical process, the convergence rate of ID empirical process can reach $O\Big(\big(\frac{\Lambda_{\mathcal{F}}(2N)}{N}\big)^{\frac{1}{1.3}}\Big)$ and it is faster than the results of the generic i.i.d. empirical process (Vapnik, 1999).

Acknowledgements

This work was supported by ARC DP Projects DP1093762 and DP0988016.

References

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

A. Bose, A. Dasgupta, and H. Rubin (2002). A Contemporary Review and Bibliography of Infinitely Divisible Distributions and Processes. *The Indian Journal of Statistics, Series A* **64**(3):763-819.

S. L. Heston (2004). Option Valuation with Infinitely Divisible Distributions. *Quantitative Finance* **4**(5):515-524.

T. Moosbrucker (2007). Copulas from Infinitely Divisible Distributions: Applications to Credit Value at Risk. *In 10th Conference of the Swiss Society for Financial Market Research.*

D. Mumford, and B. Gidas (2001). Stochastic Models for Generic Images. *Quarterly of Applied Mathematics* **59**:85-111.

P. Chainais (2007). Infinitely Divisible Cascades to Model the Statistics of Natural Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**:2105-2119.

A. van der Vaart, and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics).* Springer.

P. L. Bartlett, O. Bousquet, and S. Mendelson (2005). Local Rademacher complexities. *Annals of Statistics* **33**(4):1497-1537.

M. Mohri, and A. Rostamizadeh (2008). Rademacher Complexity Bounds for Non-i.i.d. Processes. *In NIPS 08: Neural Information Processing Systems.*

C. Zhang and D. Tao (2010). Risk Bounds for Levy Processes in the PAC-Learning Framework. *In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* **9**:948-955.

C. Houdré (2002). Remarks on Deviation Inequalities for Functions of Infinitely Divisible Random Vectors. *Annals of Probability* **30**(3):1223-1237.

N. Sauer (1972). On the Density of Families of Sets. *Journal of Combinatorial Theory, Series A* **13**(1):145-147.

K. Sato (2004). *Lévy Processes and Infinitely Divisible Distributions.* USA: Cambridge University Press.

R. N. Watteel, and R. J. Kulperger (2003). Nonparametric Estimation of the Canonical Measure for Infinitely Divisible Distributions. *Journal of Statistical Computation and Simulation* **73**(7):525-542.

R. Courant, and F. John (1974). *Introduction to Calculus and Analysis.* New York: Wiley.

C. Houdré, V. Pérez-Abreu, and D. Surgailis (1998). Interpolation, Correlation Identities, and Inequalities for Infinitely Divisible Variables. *Journal of Fourier Analysis and Applications* **4**(6):651-668.

O. Bousquet, S. Boucheron, and G. Lugosi (2004). Introduction to Statistical Learning Theory, In O. Bousquet *et al.* (ed.), *Advanced Lectures on Machine Learning*, 169-207. Springer.