RESEARCH ARTICLE

Check for updates

Exponential Inequality for the Dependent V-statistics of Bivariate Affine Functions

Richeng Zhou[1], **Weifu Li**[1] **and Liyuan Liu**[1,*]

[1] College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

Abstract

Binary functions have a wide range of applications in the fields of machine learning, statistical learning, and so on. In this paper, we investigate the exponential inequalities for the independent V-statistics of binary affine functions and obtain a universal inequality for V-statistics. Due to the typical characteristics of this kind of binary function, including symmetry and affinity, this work has great practical significance. Finally, we derive the corresponding inequalities in the context of specific similarity learning.

Keywords: V-statistics, symmetric binary affine function, exponential inequality, similarity learning.

1 Introduction

V-statistic of degree m with the function f is defined as

$$V_{m,n}(f) = \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} f\left(X_{i_1}, X_{i_2}, \cdots, X_{i_m}\right) \qquad (1)$$

where $X_1, X_2, \cdots, X_n$ are random variables taking values in a measurable space $(\mathbb{E}, \mathcal{X})$ (with $\mathbb{E}$ Polish and with distribution function ρ_i. According to (1), if we sum over different combinations of indices $(i_1, \ldots, i_m)$, the result is the U-statistics. In many applications, the analysis methods for U-statistics and V-statistics are the same (In addition to having a permutation order). Moreover, the study of non-asymptotic tail bounds and limit theorems for V-statistics and U-statistics under independent and identically distributed conditions is also very extensive [1–3]. When the observed data are no longer independent, the analysis of V-statistics and U-statistics has attracted more and more attention in the fields of statistics and probability, with most studies focusing on deriving limit theorems and the consistency of the bootstrap method, such as [4–7]. However, there are relatively few research results on the non-asymptotic concentration bounds of V-statistics and U-statistics. Exceptions include [8, 9]; in [9], exponential inequalities for m-point V-statistics were studied based on the Fourier features. In [8], concentration inequalities for U-statistics were investigated based on the Markov property.

In this paper, an exponential inequality for the V-statistics of binary affine functions [10] is derived under mixing conditions. Subsequently, this conclusion is applied to the pairwise learning scenario in machine learning, and the corresponding results are obtained.

Submitted: 04 September 2025
Accepted: 15 September 2025
Published: 18 September 2025

Vol. 1, **No.** 2, 2025.
10.62762/JNSPM.2025.502885

***Corresponding author:**
Liyuan Liu
llys@webmail.hzau.edu.cn

Citation
Zhou, R., Li, W., & Liu, L. (2025). Exponential Inequality for the Dependent V-statistics of Bivariate Affine Functions. *Journal of Numerical Simulations in Physics and Mathematics*, 1(2), 54–59.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

2 Definitions and notations

Here, we need to introduce some relevant concepts that will be used in the subsequent conclusions.

Definition 1. *A multivariate function is affine with respect to each variable X_i, which means that for any variable X_i, it satisfies the following equation:*

$$\begin{aligned} & f(X_1,\cdots,(1-\lambda)X_i^*+\lambda X_i^{**},\cdots,X_n) \\ =&(1-\lambda)f(X_1,\cdots,X_i^*,\cdots,x_n) \\ &+\lambda f(X_1,\cdots,X_i^{**},\cdots,X_n). \end{aligned}$$

The research in this paper is based on a bivariate affine function. From the properties of affine functions, we know that the function $f(X_1,X_2)$ satisfies the following inequality:

$$\begin{aligned} f(X_1,X_2)=&f((1-X_1)\cdot 0+X_1\cdot 1,x_2) \\ =&(1-X_1)f(0,X_2)+X_1f(1,X_2). \end{aligned}$$

By repeating the above decomposition for X_2, we can obtain the following decomposition formula

$$f(X_1,X_2)=\sum_{i,j=0}^{1} f_{i,j}e_i(X_1)e_j(X_2), \tag{2}$$

where $f_{i,j}=f(i,j)$, the basis functions $e_i(\cdot)$ satisfy the following conditions:

$$e_0(X)=X, e_1(X)=1-X.$$

Definition 2. *A sequence $\{X_i\}_{i\in\mathbb{N}}$ is called strong mixing (also called α-mixing) if*

$$\begin{aligned} \alpha(j):=\sup_{i\in\mathbb{N}^\star}\sup_{M_i,M_{i+j}} &|\mathbb{P}(M_i\cap M_{i+j})-\mathbb{P}(M_i)\mathbb{P}(M_{i+j})| \\ &\to 0 \text{ as } j\to\infty, \end{aligned}$$

where $M_i\in\mathcal{M}_0^i$, $M_{i+j}\in\mathcal{M}_{i+j}^\infty$, and $\mathcal{M}_a^b$ is the σ-algebra generated from random variables $X_a,\cdots,X_b$.

Denote $\theta := \theta(f) := \mathbb{E}\big[f\big(\widetilde{X}_1,\widetilde{X}_2\big)\big]$ where $\widetilde{X}_i$ are identically distributed and independent of X_i $(,i=1,2)$. f is called centered if $\theta(f)=0$, and degenerated if

$$\mathbb{E}\big[f\big(x_1,\widetilde{X}_2\big)\big]=\theta,$$

for any $x_1,x_2\in\mathcal{X}$. Using the Hoeffding decomposition, for any $x_1,x_2\in\mathcal{X}$, we have the following equation for the bivariate affine function $f(x_1,x_2)$,

$$f(x_1,x_2)=\sum_{i=1}^{2} f_1(x_i)+f_2(x_1,x_2), \tag{3}$$

where f_1,f_2 are recursively defined as

$$\begin{aligned} f_1(x_i)&=\mathbb{E}\big[f(x_1,\widetilde{X}_2)\big]-\theta, \\ f_2\left(x_1,x_2\right)&:=f(x_1,x_2)-\theta-\sum_{i=1}^{2} f_1\left(x_i\right). \end{aligned}$$

Hence, we can know that

$$f-\theta=f_1(x_1)+f_1(x_2)+f_2(x_1,x_2).$$

We denote the V-statistic generated by f_2 and sample set $\{x_i\}_{i=1}^n$ by

$$V_{2,k}(f_2)=\sum_{i_1,i_2=1}^{k} f_2(x_{i_1},x_{i_2}).$$

3 Main result and proof

Theorem 1. *Assume that $X_1,\cdots,X_n$ are n normalized samples that satisfy the following strong mixing coefficient*

$$\alpha(j)\le\gamma_1\exp(-\gamma_2 j),\quad j\ge 1,$$

where γ_1,γ_2 are two positive absolute constants. Suppose that $\|f\|\le F$ is a symmetric bivariate function. Then, we have the following inequality for $\delta>0$,

$$\mathbb{P}\left(n^{-2}\max_{1\le k\le n}|V_{2,k}\left(f_2\right)|\ge\delta\right)\le 3(A_n+M_n)e^{-n\delta}, \tag{4}$$

with

$$\begin{aligned} A_n&=\exp\left\{4\sqrt{2}F\delta C_1\left(\frac{64n\gamma_1^{1/3}}{1-\exp\left\{-\gamma_2/3\right\}}+1\right)^{1/2}\right\}, \\ M_n&=\exp\left\{12F\delta C_2(\log n)^2\right\}. \end{aligned}$$

Proof. For $p=2$, with $\{g_p\}_{p=1}^2$ defined as $g_2:=f-\theta$, so by the Hoeffding decomposition [9], we can obtain the following decomposition for binary affine functions

$$f_2(x_1,x_2)=\sum_{i,j=0}^{1} f_{i,j}\mathbb{E}(e_i)\widetilde{e}_j(x_2),$$

where $\widetilde{e}_j:=e_j-\mathbb{E}\left\{e_j(X_1)\right\}$, and

$$V_{2,k}(f_2)=\sum_{i,j=0}^{1} f_{i,j}\mathbb{E}(e_i)\left\{\sum_{t=1}^{k}\widetilde{e}_j(X_t)\right\},$$

then, for each $j = 0, 1$ and $k \in [n]$, let

$$S_{k,j} := \sum_{t=1}^{k} \widetilde{e}_j(X_t) \quad \text{and} \quad Z_j := \max_{1\le k\le n} |S_{k,j}|.$$

Since $\{X_i\}_{i=1}^n$ satisfies the α-mixing condition, $\widetilde{e}_j(X_t)$ also satisfies the α-mixing. Next, we estimate the second moment of $\max_{1\le k\le n} |V_{2,k}(f_p)|$. Let

$$T_p := n^{-p} \max_{1\le k\le n} |V_{2,k}(f_p)|.$$

Define

$$\nu = C_1 \left(n\sigma^2 + 1\right), \quad c = C_2(\log n)^2,$$

and $\sigma^2 = \sup_{j=1,2} \sigma_j^2$, with

$$\sigma_j^2 := \mathrm{Var}\left\{\widetilde{e}_j(X_1)\right\} + 2\sum_{t>1} \left|\mathrm{Cov}\left\{\widetilde{e}_j(X_1), \widetilde{e}_j(X_i)\right\}\right|.$$

Integrating the tail estimate in Corollary 24 of [11] and using Theorem 2.3 in [12] yield that, for any positive integer N, we can choosing C_2 in c such that

$$\mathbb{E}\left(Z_j^{2N}\right) \le N!(8\nu)^N + (2N)!(4c)^{2N}.$$

Let $\mu_a := \sup_{1\le t\le n}(\mathbb{E}|e_i(X_t)|^a)^{1/a} \le 1$, then we have the following from [8](Eq.(12))

$$\begin{aligned}
\mathbb{E}T_2^{2N} &= \mathbb{E}n^{-2N}\Big(\max_{1\le k\le n}\sum_{i,j=0}^{1} |f_{i,j}|\,\mathbb{E}(e_i)\,S_{k,j}\Big)^{2N} \\
&\le 4^{2N-1}n^{-2N}\mu_1^{2N}\sum_{i,j=0}^{1}|f_{i,j}|^{2N}\,\mathbb{E}\left(Z_j^{2N}\right) \\
&\le 4^{2N}n^{-2N}\mu_1^{2N}F^{2N}\left\{N!(8\nu)^N + (2N)!(4c)^{2N}\right\},
\end{aligned}$$

In the first inequality, we apply Jensen's inequality, and the second inequality holds with $\|f\| \le F$. By Stirling's approximation formula $\sqrt{2\pi}n^{n+1/2}e^{-n} \le n! \le en^{n+1/2}e^{-n}$, it holds that

$$\begin{aligned}
\{(2N)!\}^{1/2} &\le e^{1/2}(2N)^{N+1/4}e^{-N} \\
&\le e^{1/2}2^{N+1/2}N^{N+1/2}e^{-N} \le 3^N N!.
\end{aligned}$$

and we also have $(N!)^{1/2} \le N!$. Thus we have

$$\begin{aligned}
\mathbb{E}(T_2)^{\frac{2N}{2}} &\le \left(\mathbb{E}T_2^{2N}\right)^{\frac{1}{2}} \\
&\le 2^{2N}n^{-N}\mu_1^N F^N\left\{(2\sqrt{2})^N N!\nu^{N/2} + 6^N N!c^N\right\}
\end{aligned}$$

Now we control the expectation exponential transform of $T_2^{1/2}$,

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda T_2^{1/2}}\right) &= \sum_{N=0}^{\infty}\frac{\lambda^N}{N!}\mathbb{E}T_2^{N/2} \le 3\sum_{N=0}^{\infty}\frac{\lambda^{2N}}{(2N)!}\mathbb{E}T_2^{2N/2} \\
&\le 3\sum_{N=0}^{\infty}\frac{\lambda^{2N}}{(2N)!}n^{-2N}\mu_1^N F^N\left\{(8\sqrt{2})^N N!\nu^{N/2} + 24^N N!c^N\right\},
\end{aligned}$$

In the first inequality, we use only the even moments with an absolute constant 3. For the first term below, we have

$$\begin{aligned}
&\sum_{N=0}^{\infty}\lambda^{2N}\frac{N!}{(2N)!}(2\sqrt{2})^N n^{-2N}\mu_1^N F^N\nu^{N/2} \\
&\le \sum_{N=0}^{\infty}\frac{(\lambda/n)^{2N}}{N!}2^{-N}(8\sqrt{2})^N\mu_1^N F^N\nu^{N/2} \\
&= \exp\left\{\sqrt{2}(\lambda/n)^2\mu_1 F\nu^{1/2}\right\} \\
&\le \exp\left\{4\sqrt{2}(\lambda/n)^2 F\nu^{1/2}\right\}.
\end{aligned}$$

Similarly, we can derive that

$$\begin{aligned}
&\sum_{N=0}^{\infty}\lambda^{2N}\frac{N!}{(2N)!}24^N n^{-2N}\mu_1^N F^N c^N \\
&\le \sum_{N=0}^{\infty}\frac{(\lambda/n)^{2N}}{N!}2^{-N}24^N\mu_1^N F^N c^N \\
&= \exp\left\{12(\lambda/n)^2\mu_1 Fc\right\} \\
&\le \exp\left\{12(\lambda/n)^2 Fc\right\},
\end{aligned}$$

so

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda T_2^{1/2}}\right) \le\, &3\exp\left\{4\sqrt{2}(\lambda/n)^2 F\nu^{1/2}\right\} \\
&+3\exp\left\{12(\lambda/n)^2 Fc\right\}
\end{aligned}$$

Now, taking $\lambda = n\delta^{1/2}$, from Chebyshev's inequality, we can obtain

$$\mathbb{P}(T_2 \ge \delta) \le \exp\left(-\lambda\delta^{1/2}\right)\mathbb{E}\left(e^{\lambda T_2^{1/2}}\right) \le 3(A_n + M_n)e^{-n\delta},$$

where

$$A_n = 4\sqrt{2}F\delta C_1\left(n\sigma^2+1\right)^{1/2}, \quad M_n = 12F\delta C_2(\log n)^2.$$

Moreover, taking $\delta = 1$ in Theorem 3 of [13], we obtain

$$\begin{aligned}\sigma^2 =& \operatorname{Var}\left\{\tilde{e}_j\left(X_1\right)\right\}+2 \sum_{t>1}\left|\operatorname{Cov}\left\{\tilde{e}_j\left(X_1\right), \tilde{e}_j\left(X_t\right)\right\}\right| \\ \leq& 2 \sum_{t \geq 1}\left|\operatorname{Cov}\left\{\tilde{e}_j\left(X_1\right), \tilde{e}_j\left(X_t\right)\right\}\right| \\ \leq& 2\left\{\sum_{n=0}^{\infty} 8 \alpha^{1/3}(n)\right\}\left\|\tilde{e}_j\left(X_1\right)\right\|_3\left\|\tilde{e}_j\left(X_1\right)\right\|_3 \\ \leq& 64\left\{\sum_{n=0}^{\infty} \alpha^{1/3}(n)\right\}\left\|e_j\left(X_1\right)\right\|_3\left\|e_j\left(X_1\right)\right\|_3 \\ \leq& 64 \gamma_1^{1/3} \mu_3^2\left\{\sum_{n=0}^{\infty} \exp \left(-\gamma_2 n / 3\right)\right\} \\ \leq& \frac{64 \gamma_1^{1/3}}{1-\exp \left\{-\gamma_2 / 3\right\}}.\end{aligned}$$

□

With the above Theorem 1, we only need to determine f_1. In Lemma 4.3 in [9], let $p=1$; therefore, we have the following Lemma.

Lemma 1. *Given the same conditions and notation as in Theorem 1, we can obtain the following inequality.*

$$\mathbb{P}\left(n^{-1}\left|\sum_{i=1}^{n}\left(f_1(x_i)\right)\right| \geq \delta\right) \leq 6 \exp\left(-\frac{C_3 n \delta^2}{A_{1,n}+\delta M_{1,n}}\right), \tag{5}$$

where

$$\begin{aligned} A_{1,n} &= F^2\left(\frac{64\gamma_1^{1/3}}{1-\exp(-\gamma_2/3)}+\frac{(\log n)^4}{n}\right), \\ M_{1,n} &= F(\log n)^2. \end{aligned}$$

From equations (3), (4), and (5), we can derive our final result.

Theorem 2. *Assume that $X_1, \cdots, X_n$ are n normalized samples that satisfy the following strong mixing coefficient*

$$\alpha(j) \leq \gamma_1 \exp(-\gamma_2 j), \quad j \geq 1,$$

where γ_1, γ_2 are two positive absolute constants. Suppose that $\|f\| \leq F$ is a symmetric bivariate function. Then, we have the following inequality for $\delta > 0$,

$$\begin{aligned} &\mathbb{P}\left(n^{-2} \max_{1\leq k\leq n}\left|\frac{1}{n^2}\sum_{i,j=1}^{n}\left(f(x_i,x_j)-\theta\right)\right| \geq 2\delta\right) \\ \leq& 3(A_n+M_n)e^{-n\delta}+6\exp\left(-\frac{C_3 n\delta^2}{A_{1,n}+\delta M_{1,n}}\right), \end{aligned} \tag{6}$$

with

$$A_n = \exp\left\{4\sqrt{2}F\delta C_1\left(\frac{64 n\gamma_1^{1/3}}{1-\exp\{-\gamma_2/3\}}+1\right)^{1/2}\right\},$$

and

$$M_n = \exp\left\{12 F\delta C_2 (\log n)^2\right\}.$$

Proof. For simplification, we denote

$$6\exp\left(-\frac{C_3 n\delta^2}{A_{1,n}+\delta M_{1,n}}\right) = P_1, 3(A_n+M_n)e^{-n\delta} = P_2.$$

Then we can know that

$$\begin{aligned} &\mathbb{P}\left(n^{-1}\left|\sum_{i=1}^{n}\left(f_1(x_i)\right)\right| \leq \delta\right) \geq 1-P_1, \\ &\mathbb{P}\left(n^{-2}\max_{1\leq k\leq n}\left|V_{2,k}\left(f_2\right)\right| \leq \delta\right) \geq 1-P_2, \end{aligned}$$

and from equation (3)

$$\begin{aligned} &\mathbb{P}\left(n^{-2} \max_{1\leq k\leq n}\left|\frac{1}{n^2}\sum_{i,j=1}^{n}\left(f(x_i,x_j)-\theta\right)\right| \leq 2\delta\right) \\ \geq& (1-P_1)(1-P_2) \geq 1-(P_1+P_2). \end{aligned}$$

So we have the following

$$\begin{aligned} &\mathbb{P}\left(n^{-2} \max_{1\leq k\leq n}\left|\frac{1}{n^2}\sum_{i,j=1}^{n}\left(f(x_i,x_j)-\theta\right)\right| \geq 2\delta\right) \\ \leq& P_1+P_2 \\ =& 3(A_n+M_n)e^{-n\delta}+6\exp\left(-\frac{C_3 n\delta^2}{A_{1,n}+\delta M_{1,n}}\right). \end{aligned}$$

□

4 Examples and extensions

In machine learning theory, pairwise similarity learning is a crucial research scenario [21]. [14, 15] has already studied many independent and identically distributed similarity learning models, which have been applied to practical problems such as information retrieval and face recognition [16–18]. However, under mixing conditions, research is still a blank slate. For this reason, we apply the theory obtained in the previous text to similarity learning. Especially, the similarity function plays a key role in similarity learning [19, 22] which is defined as

$$s_{\boldsymbol{M}}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^{\top}\boldsymbol{M}\boldsymbol{x}_2 \tag{7}$$

where $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ is a positive definite real symmetric matrix. In this paper, we only discuss the properties of the similarity function. To ensure that the expression for similarity learning still satisfies the properties of a binary affine function, we decide to make the following transformations on $s_{\boldsymbol{M}}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and the vector $\boldsymbol{x}_i (i = 1, 2)$.

$$\begin{aligned} \boldsymbol{x}_i &\to \mathbf{X}_i = (\boldsymbol{x}_i, 0, \cdots, 0) \\ s_{\boldsymbol{M}}(\boldsymbol{x}_1, \boldsymbol{x}_2) &\to f(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{X}_1^\top \boldsymbol{M} \mathbf{X}_2 \\ &= \operatorname{diag}\{s_{\boldsymbol{M}}(\boldsymbol{x}_1, \boldsymbol{x}_2), 0, \cdots, 0\} \end{aligned} \tag{8}$$

We can verify that the $f(\mathbf{X}_1, \mathbf{X}_2)$ in equation (8) satisfies the properties of a binary affine function in matrix form

$$f(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{X}_1^\top \boldsymbol{M} \mathbf{X}_2 = \sum_{i,j=\mathbf{O}}^{\mathbf{I}} f_{i,j} e_i(\mathbf{X}_1) e_j^\top(\mathbf{X}_2)$$

where $e_{\mathbf{O}}(\mathbf{X}) = \mathbf{X}, e_{\mathbf{I}}(\mathbf{X}) = \mathbf{I} - \mathbf{X}$, the definition of $f_{i,j}$ is similar to equation (2). Therefore, we can say that the function $f(\mathbf{X}_1, \mathbf{X}_2)$ after the $s_{\boldsymbol{M}}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ transformation satisfies the property of a binary affine function. As a result, under the mixing conditions, we can obtain a conclusion similar to Theorem 2.

5 Conclusion

This paper has established the concentration estimation for symmetric binary affine functions under α-mixing. We have leveraged the decomposability of symmetric binary affine functions to factorize the original function into a product of basis functions. Building on this, we have applied the Hoeffding decomposition to transform the estimation of V-statistics into estimating inequality for maximal inequality of partial sums in a univariate form. In the future, it is of considerable interest to further extend the results to τ-mixing process [20, 23] and more general functions such as square integrable functions [24].

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Giné, E., Latała, R., & Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II* (pp. 13-38). Boston, MA: Birkhäuser Boston. [CrossRef]

[2] Arcones, M. A., & Giné, E. (1993). Limit theorems for U-processes. *The Annals of Probability*, 1494-1542.

[3] Wassily Hoeffding. (1994). Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, 409–426. [CrossRef]

[4] Ki Yoshihara. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie verw Gebiete*, 237–252. [CrossRef]

[5] Manfred Denker and Gerhard Keller. (1983). On U-statistics and v. mise'statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 64*(4), 505–522. [CrossRef]

[6] Anne Leucht and Michael H Neumann. (2013). Degenerate-and-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Annals of the Institute of Statistical Mathematics, 65*(2), 349–386. [CrossRef]

[7] Yves F Atchadé and Matias D Cattaneo. (2014). A martingale decomposition for quadratic forms of markov chains (with applications). *Stochastic Processes and their Applications, 124*(1), 646–677. [CrossRef]

[8] IS Borisov and NV Volodko. (2015). A note on exponential inequalities for the distribution tails of canonical von mises' statistics of dependent observations. *Statistics & Probability Letters, 96*, 287–291.[CrossRef]

[9] Yandi Shen, Fang Han, and Daniela Witten. (2020). Exponential inequalities for dependent V-statistics via random fourier features. *Electronic Journal of Probability, 25*. [CrossRef]

[10] Philip Schneider and David H Eberly. (2003). Geometric Tools for Computer Graphics (Morgan Kaufmann Series in Computer Graphics and Geometric Modeling). *Morgan Kaufmann Publishers*.

[11] S Boucheron, G Lugosi, and P Massart. (2013). Concentration inequalities: A nonasymptotic theory of independence. *Oxford Academic*. [CrossRef]

[12] Merlevède, F., & Peligrad, M. (2013). Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. [CrossRef]

[13] Doukhan, P. (1995). Mixing. In *Mixing: Properties and Examples* (pp. 15-23). New York, NY: Springer New York. [CrossRef]

[14] Andreas Maurer. (2008). Learning similarity with

operator-valued large-margin classifiers. *The Journal of Machine Learning Research*, *9*, 1049–1082.

[15] Kar, P., & Jain, P. (2011). Similarity-based learning via data driven embeddings. *Advances in neural information processing systems*, *24*.

[16] Hoi, S. C., Liu, W., Lyu, M. R., & Ma, W. Y. (2006, June). Learning distance metrics with contextual constraints for image retrieval. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 2, pp. 2072-2078). IEEE. [CrossRef]

[17] Guillaumin, M., Verbeek, J., & Schmid, C. (2009, September). Is that you? Metric learning approaches for face identification. In *2009 IEEE 12th international conference on computer vision* (pp. 498-505). IEEE. [CrossRef]

[18] Ying, Y., & Li, P. (2012). Distance metric learning with eigenvalue optimization. *The Journal of Machine Learning Research*, *13*(1), 1-26.

[19] Chechik, G., Sharma, V., Shalit, U., & Bengio, S. (2010). Large Scale Online Learning of Image Similarity Through Ranking. *The Journal of Machine Learning Research*, *11*, 1109-1135.

[20] Liu, L., Chen, Y., Li, W., Wang, Y., Gu, B., Zheng, F., & Chen, H. (2025). Generalization Bounds of Deep Neural Networks With τ-Mixing Samples. *IEEE Transactions on Neural Networks and Learning Systems*. [CrossRef]

[21] Liu, L., Chen, H., Xiao, C., & Li, W. (2025). The consistency analysis of gradient learning under independent covariate shift. *Neurocomputing*, *635*, 129883. [CrossRef]

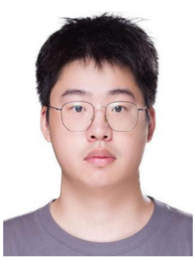
[22] Cao, Q., Guo, Z. C., & Ying, Y. (2016). Generalization bounds for metric and similarity learning. *Machine Learning*, *102*(1), 115-132. [CrossRef]

[23] Dedecker, J., & Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probability Theory and Related Fields*, *132*(2), 203-236. [CrossRef]

[24] Duchemin, Q., De Castro, Y., & Lacour, C. (2023). Concentration inequality for U-statistics of order two for uniformly ergodic Markov chains. *Bernoulli*, *29*(2), 929-956. [CrossRef]

Richeng Zhou received the B.S. degree from Harbin Insititute of Technology, Haerbin, China, in 2021. He is currently pursuing the M.S. degree with the College of Informatics, Huazhong Agricultural University, Wuhan 430070, China. (Email: zhourc@webmail.hzau.edu.cn)

Weifu Li received the B.S., M.S., and Ph.D. degrees from Hubei University, Wuhan, China, in 2014, 2017, and 2020, respectively. From September 2019 to July 2020, he was a joint Student with the Research Center for Smart Vehicles, Toyota Technological Institute, Nagoya, Japan. He is currently an Associate Professor with the Department of Mathematics and Statistics, College of Informatics, Huazhong Agricultural University, Wuhan. (Email: liweifu@mail.hzau.edu.cn)



Liyuan Liu received the B.S. degree from Hubei University of Technology, Wuhan, China, in 2021. He is currently pursuing the Ph.D. degree with the College of Informatics, Huazhong Agricultural University, Wuhan, China. His research interests include machine learning and statistical learning theory. (Email: llys@webmail.hzau.edu.cn)