

Some Theoretical Results of Learning Theory Based on Intuitionistic Fuzzy Random Samples

Zhiming Zhang*

College of Mathematics and Computer Science, Hebei University, Baoding 071002, Hebei
Province, PR China

ABSTRACT

In this paper, we discuss the statistical learning theory based on intuitionistic fuzzy random sample. First of all, we introduce the definition of intuitionistic fuzzy number intuitionistic fuzzy random variables. Secondly, we give some properties of this concept. Thirdly, the definitions of intuitionistic fuzzy empirical risk functional, intuitionistic fuzzy expected risk functional, and intuitionistic fuzzy empirical risk minimization principle are presented. Finally, we prove the key theorem of intuitionistic fuzzy random sample and obtain the rate of uniform convergence of learning process based on the intuitionistic fuzzy random sample.

Keywords: Intuitionistic fuzzy numbers; Intuitionistic fuzzy random variables; Intuitionistic fuzzy empirical risk minimization principle; The key theorem; The bounds of the rate of convergence

1. INTRODUCTION

Statistical learning theory (SLT, for short), put forward in 1960s and completely founded by Vapnik et al. in 1990s. [46-48], has become an interesting and good law that supports the development of small samples statistical learning. The SLT has become the fastest growing discipline in machine learning in the late 1990's. Its essence was to make the learning machines work effectively with the limited samples and then improve the generalization abilities of the learning machines. By doing this, we establish a meaningful theoretical framework for statistical learning based on small data samples. Meanwhile, SLT gave rise to a new category of general learning algorithms, which we called the Support Vector Machine (SVM, for short). Currently, the SLT and SVM constitute interesting research avenues in machine learning [6, 7, 10, 11, 12, 20, 28, 29, 30, 31, 35, 36, 39, 41, 43, 44, 45, 49, 50, 52, 53, 56].

The SLT covers four main parts [6]: (1) the learning process that minimizes the necessary and sufficient conditions for the consistency of the empirical risk, which is referred to as the key theorem of learning theory; (2) the scope of generalization; (3) the structural risk minimization principle; (4) the support vector machine (SVM) algorithm that implements the structure risk minimization principle. The key theorem of learning theory and the boundary of convergence of the learning process are the two most important foundations for further research on related topics, such as the structural risk minimization and the SVM.

Although the SLT has reached a level of high maturity, there are still some problems to be solved, for example, the development of SLT and SVM is based on probability space and real-valued random sample (real-valued random variables). In the real world, there are many non-probability space (such as the Sugeno measure space [51], credibility measure space [38]) and non-real random samples (such as fuzzy sample [55], fuzzy random sample [32], and fuzzy complex random samples [26]). In order to solve these problems, the expansions of the SLT in non-probabilistic space and non-real random samples have become an urgent need to move forward. Some studies have been achieved along this line [21-27, 39, 42, 43]. For example, Ha et al. [21-23] studied the key theorem and the generalized boundary on

* Tel.: +86 312 5079660; fax: +86 312 5079660.
E-mail address: zhimingzhang@ymail.com.

Sugeno measure probability space and quasi-probability measure space; Tian et al. [24,26] investigated the construction of statistical learning theory based on fuzzy random samples and fuzzy complex random samples; Lin et al. [36] constructed fuzzy support vector machine based on random sample.

Intuitionistic (IF, for short) set, originated by Atanassov [1,5], is an important concept used to cope with imperfect and/or imprecise information. It is an intuitively straightforward extension of Zadeh's fuzzy sets [55]: while a fuzzy set gives a degree to which an element belongs to a set, an intuitionistic fuzzy set gives both a membership degree and a non-membership degree. The membership and non-membership values induce an indeterminacy index, which models the hesitancy of deciding the degree to which an object satisfies a particular property. As the basis for the study of IF set theory, many operations and relations over IF sets were introduced [1-5]. Many concepts in fuzzy set theory were also extended to IF set theory, such as IF relations, intuitionistic L-fuzzy sets, IF implication, IF logics, and the degree of similarity between IF sets, etc., [4,9,13,15,16,33,34,57]. For a further study for the structure of IF sets, construction theorems of IF sets, IF topology and the axiomatic characterization of IF sets have been investigated [8,14,18,40]. Recently, IF set theory has been successfully applied in decision analysis and pattern recognition (see, e.g., [19,37,54]).

However, the study for the combination of IF set theory and statistical learning theory is still blank. This paper discusses the statistical learning theory based on intuitionistic fuzzy random samples by combining intuitionistic fuzzy analysis and SLT, revisits the major parts of the SLT and establishes some ground material for further development of classifiers such as support vector machines. This paper is organized as follows: Section 2 introduces some basic definitions, elaborates on a number of properties of intuitionistic fuzzy random variables. In Section 3, we prove the key theorem of learning theory based on intuitionistic fuzzy random samples. In the sequel, in Section 4, the bounds of the rate of uniform convergence of learning process based on intuitionistic fuzzy random samples are discussed. The final section offers the conclusions and brings prospects of potential future developments.

2. PRELIMINARIES

Throughout this paper, we let $(\Omega, \mathcal{A}, P)$ be a probability measure space and $\mathbb{R}$ be the real numbers field.

Definition 2.1 [32]. A fuzzy number is a fuzzy set $X : \mathbb{R} \rightarrow [0,1]$ with the following properties:

- (1) X is normal, i.e., there exists $x \in \mathbb{R}$ such that $X(x) = 1$;
- (2) X is upper semi-continuous;
- (3) $\text{supp } X = \text{cl}\{x \in \mathbb{R} : X(x) > 0\}$ is compact;
- (4) X is a convex fuzzy set, i.e., $X(\lambda x + (1-\lambda)y) \geq \min(X(x), X(y))$, for $\forall x, y \in \mathbb{R}$ and $\lambda \in [0,1]$.

Let $F(\mathbb{R})$ be the family of all fuzzy numbers on $\mathbb{R}$. For a fuzzy set X , if we define

$$(X)_r = \begin{cases} \{x : X(x) \geq r\}, & 0 < r \leq 1 \\ \text{supp } X, & r = 0 \end{cases},$$

then it follows that X is a fuzzy number if and only if $(X)_1 \neq \emptyset$ and $(X)_r$ is a closed bounded interval for each $r \in [0,1]$. Therefore, a fuzzy number X is completely determined by the intervals $(X)_r$.

Let $X, Y \in F(\mathbb{R})$, $\forall r \in [0,1]$, $(X)_r = [(X)_r^-, (X)_r^+]$ and $(Y)_r = [(Y)_r^-, (Y)_r^+]$. Define

$$X \pm Y = \bigcup_{r \in [0,1]} \left\{ \left[(X)_r^- \pm (Y)_r^-, (X)_r^+ \pm (Y)_r^+ \right] \right\}.$$

Let us introduce a partially ordered relation on $F(\mathbb{R})$ as follow: for any $X, Y \in F(\mathbb{R})$, we say $X \leq Y$, iff $\forall r \in [0,1]$, $(X)_r \leq (Y)_r$ i.e., $(X)_r^- \leq (Y)_r^-$ and $(X)_r^+ \leq (Y)_r^+$.

Definition 2.2 [32]. A fuzzy number valued function $\xi : \Omega \rightarrow F(\mathbb{R})$ is called fuzzy random variable if for every closed subset B of $\mathbb{R}$, the fuzzy set $\xi^{-1}(B)$ is measurable when considered as a function from Ω to $[0,1]$, where $\xi^{-1}(B)$ denotes the fuzzy subset of Ω defined by $\xi^{-1}(B)(\omega) = \sup_{x \in B} \xi(\omega)(x)$ for every $\omega \in \Omega$.

Definition 2.3 [32]. A fuzzy random variable $\xi(\omega) = \left[\xi_r^-(\omega), \xi_r^+(\omega) \right] | 0 < r \leq 1$ is called integrable if for each $r \in [0,1]$, $\xi_r^-(\omega)$ and $\xi_r^+(\omega)$ are integrable, or equivalently $\int \|\xi\| dP < \infty$, where $\|\xi\| = \max \left\{ \left| \xi_0^- \right|, \left| \xi_0^+ \right| \right\}$. In this case, the expected value of ξ is defined in the following manner

$$E\xi = \int \xi dP = \left\{ \left[\int \xi_r^- dP, \int \xi_r^+ dP \right] | 0 < r \leq 1 \right\}.$$

Definition 2.4 [1]. Let U be a nonempty set. An intuitionistic fuzzy set A in U is an object having the form $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle | x \in U \}$,

where $\mu_A : U \rightarrow [0,1]$ and $\nu_A : U \rightarrow [0,1]$ satisfy $0 \leq \mu_A + \nu_A \leq 1$ for all $x \in U$, and $\mu_A(x)$ and $\nu(x)$ are, respectively, called the degree of membership and the degree of non-membership of the element $x \in U$ to A . The complement of an IF set A is denoted by $A^c = \{ \langle x, \nu(x), \mu_A(x) \rangle | x \in U \}$.

Obviously, every fuzzy set $A = \{ \langle x, A(x) \rangle | x \in U \} = \{ \langle x, \mu_A(x) \rangle | x \in U \}$ can be identified with the IF set of the form $\{ \langle x, \mu_A(x), 1 - \mu_A(x) \rangle | x \in U \}$ and is thus an IF set.

We introduce some basic operations about IF sets as follows [1–5,17,33]:

Let A and B be two IF sets, then

$A \subseteq B$ iff $\mu_A(x) \leq \mu_B(x)$ and $\nu_A(x) \leq \nu_B(x)$ for all $x \in U$,

$A \supseteq B$ iff $B \subseteq A$,

$A = B$ iff $A \subseteq B$ and $B \subseteq A$, i.e., $\mu_A(x) = \mu_B(x)$ and $\nu_A(x) = \nu_B(x)$ for any $x \in U$,

$A \cap B = \{ \langle x, \mu_A(x) \wedge \mu_B(x), \nu_A(x) \vee \nu_B(x) \rangle | x \in U \}$,

$A \cup B = \{ \langle x, \mu_A(x) \vee \mu_B(x), \nu_A(x) \wedge \nu_B(x) \rangle | x \in U \}$,

$\forall \alpha, \beta \in [0,1], \alpha + \beta \leq 1, (\alpha, \beta)A = \{ \langle x, \alpha \wedge \mu_A(x), \beta \vee \nu_A(x) \rangle | x \in U \}$.

Definition 2.5 [58]. Let A be an IF set and $\alpha, \beta \in [0,1]$ with $\alpha + \beta \leq 1$, the (α, β) -level cut set of A , denoted by A_{α}^{β} , is defined as follows:

$$A_{\alpha}^{\beta} = \{ x \in U | \mu_A(x) \geq \alpha, \nu_A(x) \leq \beta \}$$

$A_{\alpha} = \{ x \in U | \mu_A(x) \geq \alpha \}$ and $A_{\alpha+} = \{ x \in U | \mu_A(x) > \alpha \}$ are, respectively, called the α -level cut set and the strong α -level cut set of membership generated by A . And

124 $A^\beta = \{x \in U \mid \nu_A(x) \leq \beta\}$ and $A^{\beta+} = \{x \in U \mid \nu_A(x) < \beta\}$ are, respectively, referred to as the
 125 β -level cut set and the strong β -level cut set of non-membership generated by A .

126 **Proposition 2.1.** Let $A \in IF(U)$, $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$. Then

127
$$A = \bigcup (\alpha, \beta) A_\alpha^\beta = \bigcup (\alpha, \beta) A_{\alpha+}^\beta = \bigcup (\alpha, \beta) A_\alpha^{\beta+} = \bigcup (\alpha, \beta) A_{\alpha+}^{\beta+}.$$

128 **Proof.** We have only to prove that the equation $A = \bigcup (\alpha, \beta) A_\alpha^\beta$ holds for any $x \in U$. The
 129 other equations can be proved in a similar way. On one hand, we have

130
$$\begin{aligned} \mu_{\bigcup (\alpha, \beta) A_\alpha^\beta}(x) &= \vee \mu_{(\alpha, \beta) A_\alpha^\beta}(x) = \vee (\alpha \wedge A_\alpha^\beta(x)) \\ &= \left(\bigvee_{0 \leq \alpha \leq \mu_A(x)} (\alpha \wedge A_\alpha^\beta(x)) \right) \vee \left(\bigvee_{\mu_A(x) < \alpha \leq 1} (\alpha \wedge A_\alpha^\beta(x)) \right) \\ &= \bigvee_{0 \leq \alpha \leq \mu_A(x)} \alpha = \mu_A(x). \end{aligned}$$

131 On the other hand, we have

132
$$\begin{aligned} \nu_{\bigcup (\alpha, \beta) A_\alpha^\beta}(x) &= \wedge \nu_{(\alpha, \beta) A_\alpha^\beta}(x) = \wedge (\beta \vee (1 - A_\alpha^\beta(x))) \\ &= \left(\bigwedge_{0 \leq \beta \leq \nu_A(x)} (\beta \vee (1 - A_\alpha^\beta(x))) \right) \wedge \left(\bigwedge_{\nu_A(x) < \beta \leq 1} (\beta \vee (1 - A_\alpha^\beta(x))) \right) \\ &= \bigwedge_{\nu_A(x) < \beta \leq 1} \beta = \nu_A(x). \end{aligned}$$

133 Thus $A = \bigcup (\alpha, \beta) A_\alpha^\beta$.

134 **Definition 2.6** [59]. The IF set $A = \{\langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in R\}$ is called an IF number if and only
 135 if μ_A and ν_A^c are two fuzzy numbers, where $\nu_A^c = 1 - \nu_A$.

136 The family of all intuitionistic fuzzy numbers is denoted by $IF(R)$.

137 **Proposition 2.2.** Let $A \in IF(R)$. Then the membership function $\mu_A(x)$ and non-membership
 138 function $\nu_A(x)$ have the following properties:

- 139 (1) $\mu_A : R \rightarrow [0, 1]$ is an upper semi-continuous function and $\mu_A : R \rightarrow [0, 1]$ is a lower semi-
 140 continuous function;
 141 (2) $\mu_A(x) = 0, \forall x \in (-\infty, c] \cup [d, +\infty)$, $\nu_A(x) = 1, \forall x \in (-\infty, e] \cup [f, +\infty)$;
 142 (3) $\mu_A(x) = 1, \nu_A(x) = 0, \forall x \in [a, b]$;
 143 (4) $\mu_A(x)$ is strictly monotonously increasing on $[c, a]$ and strictly monotonously decreasing
 144 on $[b, d]$; $\nu_A(x)$ is strictly monotonously decreasing on $[e, a]$ and strictly monotonously
 145 increasing on $[b, f]$.

146 **Definition 2.7.** Some Operations on $IF(R)$ are defined as follows:

147 $\forall A, B \in IF(R),$

148
$$A + B = \left\{ \left\langle z, \sup_{x+y=z} (\mu_A(x) \wedge \mu_B(y)), \inf_{x+y=z} (\nu_A(x) \vee \nu_B(y)) \right\rangle \mid x, y \in U \right\};$$

149
$$\rho A = \left\{ \left\langle z, \mu_A\left(\frac{z}{\rho}\right), \nu_A\left(\frac{z}{\rho}\right) \right\rangle \mid z = \rho x, x \in U \right\}, \rho \neq 0;$$

150 If $\rho = 0$, then

151
$$\mu_{0A}(z) = \begin{cases} \sup \{\mu_A(x) \mid x \in U\}, & z = 0 \\ 0, & z \neq 0 \end{cases}, \nu_{0A}(z) = \begin{cases} \inf \{\nu_A(x) \mid x \in U\}, & z = 0 \\ 1, & z \neq 0 \end{cases}.$$

152 **Proposition 2.3.** Let $A, B \in IF(\mathbb{R})$. And (α, β) -level cut set of A and B are denoted by

153 $A_\alpha^\beta = [\mu_A^L, \mu_A^R, \nu_A^L, \nu_A^R, (\alpha, \beta)]$, $B_\alpha^\beta = [\mu_B^L, \mu_B^R, \nu_B^L, \nu_B^R, (\alpha, \beta)]$. Then we have

154 $A + B = \left\{ [\mu_A^L + \mu_B^L, \mu_A^R + \mu_B^R, \nu_A^L + \nu_B^L, \nu_A^R + \nu_B^R, (\alpha, \beta)] \mid \alpha, \beta \in [0, 1], \alpha + \beta \leq 1 \right\}$,

155 $\rho A = \begin{cases} [\rho \mu_A^L, \rho \mu_A^R, \rho \nu_A^L, \rho \nu_A^R, (\alpha, \beta)] \mid \alpha, \beta \in [0, 1], \alpha + \beta \leq 1, & \text{if } \rho \geq 0 \\ [\rho \mu_A^R, \rho \mu_A^L, \rho \nu_A^R, \rho \nu_A^L, (\alpha, \beta)] \mid \alpha, \beta \in [0, 1], \alpha + \beta \leq 1, & \text{if } \rho < 0 \end{cases}$.

156 **Definition 2.8** [58]. Let $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in U \}$, $B = \{ \langle x, \mu_B(x), \nu_B(x) \rangle \mid x \in U \} \in IF(\mathbb{R})$.

157 Define $A \leq B$ if $\mu_A \leq \mu_B$ and $\nu_A^c \leq \nu_B^c$; $A = B$ if $A \leq B$ and $B \leq A$; $A < B$ if $A \leq B$ and $A \neq B$.

158 Obviously, $(IF(\mathbb{R}), \leq)$ is a partially ordered set.

159 **Definition 2.9.** Let $S \subseteq IF(\mathbb{R})$ and $M \in IF(\mathbb{R})$. We say that M is the supremum of the set S if
160 the following two conditions are satisfied:

161 (1) for any $A \in S$, $A \leq M$, i.e., M is the upper bound of the set S ;

162 (2) for any the upper bound N of S , $M \leq N$.

163 Similarly, we can define the infimum of the set S .

164 Let $K(\mathbb{R})$ be a family of all the nonempty compact convex subsets of $\mathbb{R}$. If $A, B \in K(\mathbb{R})$,
165 then the Hausdorff metric is defined by

166 $d_H(A, B) = \max \left\{ \sup_{x \in A} \inf_{y \in B} d(x, y), \sup_{y \in B} \inf_{x \in A} d(x, y) \right\}$,

167 where $d(x, y)$ denotes the distance between two real numbers x and y .

168 Let us define a consistent Hausdorff metric in $IF(\mathbb{R})$ to be in the following form

$$\begin{aligned} d(A, B) &= \sup_{\substack{\alpha, \beta \in [0, 1] \\ \alpha + \beta \leq 1}} \max \{ d_H(A_\alpha, B_\alpha), d_H(A^\beta, B^\beta) \} \\ &= \max \{ d(\mu_A, \mu_B), d(\nu_A, \nu_B) \} \\ &= \max \{ d(\mu_A, \mu_B), d(\nu_A^c, \nu_B^c) \} \end{aligned}$$

170 where $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in U \}$, $B = \{ \langle x, \mu_B(x), \nu_B(x) \rangle \mid x \in U \} \in IF(\mathbb{R})$.

171 **Proposition 2.4.** If $A, B, C \in IF(\mathbb{R})$, then $d(A, B) + d(B, C) \geq d(A, C)$.

172 **Proposition 2.5.** If $A, B, C \in IF(\mathbb{R})$ and $A \geq B \geq C$,
173 then $d(A, C) \geq d(A, B)$ and $d(A, C) \geq d(B, C)$.

174 **Definition 2.10.** Let $(\Omega, \mathcal{A}, P)$ be a probability measure space. An intuitionistic fuzzy number
175 valued mapping $\xi: \Omega \rightarrow IF(\mathbb{R})$, $\omega \rightarrow \xi(\omega) = \{ \langle x, \mu_{\xi(\omega)}(x), \nu_{\xi(\omega)}(x) \rangle \mid x \in \mathbb{R} \}$ is called an
176 intuitionistic fuzzy random variable if μ_ξ and ν_ξ^c are two fuzzy random variables defined on
177 $(\Omega, \mathcal{A}, P)$, where $\mu_\xi: \omega \rightarrow \mu_{\xi(\omega)}$, $\nu_\xi: \omega \rightarrow \nu_{\xi(\omega)}$ and $\nu_\xi^c = 1 - \nu_\xi$.

178 **Definition 2.11** Let $\xi(\omega) = \{ \langle x, \mu_{\xi(\omega)}(x), \nu_{\xi(\omega)}(x) \rangle \mid x \in \mathbb{R} \}$ be the intuitionistic fuzzy random
179 variable defined on $(\Omega, \mathcal{A})$. We called

180 $F(Z) = P\{ \xi(\omega) \leq Z \} = P\{ \mu_{\xi(\omega)} \leq \mu_Z, \nu_{\xi(\omega)}^c \geq \nu_Z^c \}$

181 as the distribution function of ξ , where $Z = \{ \langle x, \mu_Z(x), \nu_Z(x) \rangle \mid x \in \mathbb{R} \} \in IF(\mathbb{R})$.

Definition 2.12. We call $\xi(\omega) = \left\{ \langle x, \mu_{\xi(\omega)}(x), \nu_{\xi(\omega)}(x) \rangle \mid x \in \mathbb{R} \right\}$ integrable if μ_{ξ} and ν_{ξ}^c are integrable. In this case, we define the mathematical expectation of ξ as the following manner

$$E(\xi) = \int \xi dP = \left\{ \langle x, E(\mu_{\xi})(x), E(\nu_{\xi})(x) \rangle \mid x \in \mathbb{R} \right\} \\ = \left\{ \left[\int \mu_{\xi}^L dP, \int \mu_{\xi}^R dP, \int \nu_{\xi}^L dP, \int \nu_{\xi}^R dP, (\alpha, \beta) \right] \mid \alpha, \beta \in [0, 1], \alpha + \beta \leq 1 \right\},$$

where $E(\nu_{\xi}) = 1 - E(\nu_{\xi}^c)$.

Proposition 2.6. Let ξ be an intuitionistic fuzzy random variable, then the following equalities

1. $\{E(\xi)\}_{\alpha}^{\beta} = E(\xi_{\alpha}^{\beta}) = E(\xi_{\alpha}) \cap E(\xi^{\beta})$ for $0 < r \leq 1$;
 2. $E(c\xi) = cE\xi$, whenever $c \in \mathbb{R}$;
 3. $E(\xi_1 \pm \xi_2) = E\xi_1 \pm E\xi_2$
- hold.

Definition 2.13. Assume that $\xi_n = \left\{ \langle x, \mu_{\xi_n}(x), \nu_{\xi_n}(x) \rangle \mid x \in U \right\}, n = 1, 2, \dots$ is a collection of intuitionistic fuzzy random variables.

(1) If the fuzzy random variables sequence $\mu_{\xi_n}, n = 1, 2, \dots$ and $\nu_{\xi_n}^c, n = 1, 2, \dots$ are respectively mutually independent, then $\xi_n, n = 1, 2, \dots$ is called a collection of mutually independent intuitionistic fuzzy random variables.

(2) If $\mu_{\xi_n}, n = 1, 2, \dots$ and $\nu_{\xi_n}^c, n = 1, 2, \dots$ are two sequences of identically distributed fuzzy random vectors, then $\xi_n, n = 1, 2, \dots$ is referred to as a sequence of identically distributed intuitionistic fuzzy random variables.

Definition 2.14. Let $\xi_n, n = 1, 2, \dots$ be a sequence of intuitionistic fuzzy random variables and let ξ be an intuitionistic fuzzy random variable (or an intuitionistic fuzzy number). If $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left\{ \omega \mid d(\xi_n(\omega), \xi(\omega)) > \varepsilon \right\} = 0,$$

then ξ_n converge in probability P to ξ , denoted by

$$\xi_n \xrightarrow{P} \xi, n \rightarrow \infty, \text{ or } \lim_{n \rightarrow \infty} \xi_n = \xi.$$

Theorem 2.1. Suppose that $\xi_n = \left\{ \langle x, \mu_{\xi_n}(x), \nu_{\xi_n}(x) \rangle \mid x \in U \right\}, n = 1, 2, \dots$ is a sequence of intuitionistic fuzzy random variables and $\xi = \left\{ \langle x, \mu_{\xi}(x), \nu_{\xi}(x) \rangle \mid x \in U \right\}$ is an intuitionistic fuzzy random variable. Then, $\xi_n \xrightarrow{P} \xi$ if and only if $\mu_{\xi_n} \xrightarrow{P} \mu_{\xi}$ and $\nu_{\xi_n} \xrightarrow{P} \nu_{\xi}$.

Proof. By the definition of Hausdorff metric in $IF(\mathbb{R})$, we have

$$d(\mu_{\xi_n}, \mu_{\xi}) \leq d(\xi_n, \xi), d(\nu_{\xi_n}, \nu_{\xi}) \leq d(\xi_n, \xi), d(\xi_n, \xi) \leq d(\mu_{\xi_n}, \mu_{\xi}) + d(\nu_{\xi_n}, \nu_{\xi}).$$

Hence,

$$P\{d(\xi_n, \xi) > \varepsilon\} \leq P\{d(\mu_{\xi_n}, \mu_{\xi}) + d(\nu_{\xi_n}, \nu_{\xi}) > \varepsilon\} \leq P\left\{d(\mu_{\xi_n}, \mu_{\xi}) > \frac{\varepsilon}{2}\right\} + P\left\{d(\nu_{\xi_n}, \nu_{\xi}) > \frac{\varepsilon}{2}\right\}.$$

The theorem is easily proven.

Theorem 2.2 (The strong law of large numbers of fuzzy random variables) [32].

Let $\{\xi_n, n = 1, 2, \dots\}$ be a sequence of independent and identically distributed fuzzy random variables with $E\|\xi_1\| < \infty$. Then we have

$$\frac{1}{n} \sum_{j=1}^n \xi_j \xrightarrow{P} E\xi_1, n \rightarrow \infty.$$

Theorem 2.3 (The strong law of large numbers of intuitionistic fuzzy random variables). Let $\{\xi_n, n=1,2,\dots\}$ be a sequence of independent and identically distributed intuitionistic fuzzy random variables with $E\|\mu_{\xi_1}\| < \infty$ and $E\|v_{\xi_1}\| < \infty$. Then

$$\frac{1}{n} \sum_{j=1}^n \xi_j \xrightarrow{P} E\xi_1, n \rightarrow \infty.$$

Proof. It follows from Theorem 2.2 and Proposition 2.3.

222

223 3. THE KEY THEOREM OF LEARNING THEORY BASED ON INTUITIONISTIC 224 FUZZY RANDOM SAMPLES

225

226 Let $\xi_j, j=1,2,\dots,l$ be a sequence of independent and identically distributed intuitionistic fuzzy
227 random samples whose distribution is given as $F(\xi)$.

228 **Definition 3.1.** We call $R_{if}(\alpha) = E[Q(\xi, \alpha)] = \int Q(\xi, \alpha) dP, \alpha \in \Lambda$, where Λ is an index set
229 and $Q(\xi, \alpha)$ denotes the loss functional, the expected risk functional on the basis of
230 intuitionistic fuzzy random samples. It could be considered as the intuitionistic fuzzy
231 expected risk functional.

232 **Definition 3.2.** $R_{ifemp}(\alpha) = \frac{1}{l} \sum_{j=1}^l Q(\xi_j, \alpha), \alpha \in \Lambda$ is called the intuitionistic fuzzy empirical risk
233 functional.

234 Let the risk functional obtain its minimum at $Q(\xi, \alpha_0)$ and the empirical risk functional
235 obtain its minimum at $Q(\xi, \alpha_l)$.

236 **Definition 3.3.** We minimize $R_{ifemp}(\alpha)$ to replace $R_{if}(\alpha)$ and refer to the function $Q(\xi, \alpha_l)$
237 as an approximation to the function $Q(\xi, \alpha_0)$. We call this principle the intuitionistic fuzzy
238 empirical risk minimization principle.

239 **Definition 3.4.** If the following two sequences converge in probability to the same limit:

$$240 R_{if}(\alpha_l) \xrightarrow[l \rightarrow \infty]{P} \inf_{\alpha \in \Lambda} R_{if}(\alpha),$$

$$241 R_{ifemp}(\alpha_l) \xrightarrow[l \rightarrow \infty]{P} \inf_{\alpha \in \Lambda} R_{if}(\alpha),$$

242 then the intuitionistic fuzzy empirical risk minimization principle is consistent for the set of
243 functions $Q(\xi, \alpha), \alpha \in \Lambda$ and for the distribution functional $F(\xi)$

244 **Definition 3.5.** If for any nonempty subset $\Lambda(C), C \in IF(\mathbb{R})$ of this set of functions such that

$$245 \Lambda(C) = \left\{ \alpha : \int Q(\xi, \alpha) dP \geq C \right\} \text{ the convergence}$$

$$246 \inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha) \xrightarrow[l \rightarrow \infty]{P} \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha),$$

247 then the intuitionistic fuzzy empirical risk minimization principle is strictly consistent for the
248 set of function $Q(\xi, \alpha), \alpha \in \Lambda$ and the distribution function $F(\xi)$.

249 **Remark 3.1.** We only consider the sequence of intuitionistic fuzzy random variables that are
250 partially ordered " $\leq$ ". We denote by $d\left(\left(R_{if}(\alpha), R_{ifemp}(\alpha)\right)_{\geq}\right)$ the Hausdorff metric
251 when $R_{ifemp}(\alpha) \leq R_{if}(\alpha)$.

Theorem 3.1 (The key theorem of learning theory based on intuitionistic fuzzy random samples). Suppose that there exist two intuitionistic fuzzy numbers W_1 and W_2 , such that for all functions in the set $\mathcal{Q}(\xi, \alpha)$, $\alpha \in \Lambda$, and for a given distributed function $F(\xi)$, the inequalities $W_1 \leq \int \mathcal{Q}(\xi, \alpha) dP \leq W_2$ hold true. Then the sufficient and necessary conditions for the strict consistency of IFERM principle is that the convergence

$$P \left\{ \sup_{\alpha \in \Lambda} d \left(\left(R_{if}(\alpha), R_{iferm}(\alpha) \right)_{\geq} \right) > \varepsilon \right\} \xrightarrow{l \rightarrow \infty} 0$$

holds for any $\varepsilon > 0$.

Proof. Necessary: By Definition 3.5 for the set $\Lambda(C) = \{\alpha : R_{if}(\alpha) \geq C\}$, we have

$$\inf_{\alpha \in \Lambda(C)} R_{iferm}(\alpha) \xrightarrow{l \rightarrow \infty} \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha) \quad (3.1)$$

We denote by A the event of the form $\sup_{\alpha \in \Lambda} d \left(\left(R_{if}(\alpha), R_{iferm}(\alpha) \right)_{\geq} \right) > \varepsilon$.

Suppose that A holds. Then there exists $\alpha^* \in \Lambda(C)$ such that $d \left(\left(R_{if}(\alpha^*), R_{iferm}(\alpha^*) \right)_{\geq} \right) > \varepsilon$. We

can find $W_1 \leq a_k \leq W_2$, such that $\alpha^* \in \Lambda(a_k)$ and $d \left(\left(R_{if}(\alpha^*), a_k \right)_{\geq} \right) < \frac{\varepsilon}{2}$. Then for these $\Lambda(a_k)$ the inequalities

$$d \left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha) \right)_{\geq} \right) < \frac{\varepsilon}{2} \text{ and } R_{iferm}(\alpha^*) \geq \inf_{\alpha \in \Lambda(a_k)} R_{iferm}(\alpha)$$

hold true.

Therefore

$$\begin{aligned} & d \left(\left(\inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(a_k)} R_{iferm}(\alpha) \right)_{\geq} \right) \\ & \geq d \left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(a_k)} R_{iferm}(\alpha) \right)_{\geq} \right) - d \left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha) \right)_{\geq} \right) > \frac{\varepsilon}{2}. \end{aligned}$$

According to (3.1), we have

$$P \left\{ d \left(\left(\inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(a_k)} R_{iferm}(\alpha) \right)_{\geq} \right) > \frac{\varepsilon}{2} \right\} \xrightarrow{l \rightarrow \infty} 0.$$

Denote by T_k the event $d \left(\left(\inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(a_k)} R_{iferm}(\alpha) \right)_{\geq} \right) > \frac{\varepsilon}{2}$, then $A \subseteq \bigcup_k T_k$,

$$P(A) \leq P \left(\bigcup_k T_k \right) \xrightarrow{l \rightarrow \infty} 0.$$

Hence

$$P \left\{ \sup_{\alpha \in \Lambda} d \left(\left(R_{if}(\alpha), R_{iferm}(\alpha) \right)_{\geq} \right) > \varepsilon \right\} \xrightarrow{l \rightarrow \infty} 0 \quad (3.2)$$

Sufficiency: We denote by A the event $\left\{ \omega : d \left(\left(\inf_{\alpha \in \Lambda(C)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{iferm}(\alpha) \right)_{\geq} \right) > \varepsilon \right\}$

$$A_1 = \left\{ \omega : d \left(\left(\inf_{\alpha \in \Lambda(C)} R_{iferm}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha) \right)_{\geq} \right) > \varepsilon \right\}$$

$$A_2 = \left\{ \omega : d \left(\left(\inf_{\alpha \in \Lambda(C)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{iferm}(\alpha) \right)_{\geq} \right) > \varepsilon \right\}.$$

Then $A = A_1 \cup A_2$, and $P(A) \leq P(A_1) + P(A_2)$.

Suppose that the event A_1 occurs, we can find $\mathcal{Q}(\xi, \alpha^*), \alpha^* \in \Lambda(C)$, such that

$$d\left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)\right)_{\geq}\right) < \frac{\varepsilon}{2} \text{ and } R_{ifemp}(\alpha^*) \geq \inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha).$$

$$\text{By } d\left(\left(\inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)\right)_{\geq}\right) > \varepsilon, \text{ we have}$$

$$d\left(\left(R_{ifemp}(\alpha^*), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)\right)_{\geq}\right) > \varepsilon.$$

Therefore

$$d\left(\left(R_{ifemp}(\alpha^*), R_{if}(\alpha^*)\right)_{\geq}\right) > \frac{\varepsilon}{2}.$$

In virtue of the monotonicity of probability and the strong law of large numbers of intuitionistic fuzzy random variables, we obtain

$$P(A_1) \leq P\left\{d\left(\left(R_{ifemp}(\alpha^*), R_{if}(\alpha^*)\right)_{\geq}\right) > \frac{\varepsilon}{2}\right\} \xrightarrow{l \rightarrow \infty} 0.$$

On the other hand, if the event A_2 takes place, we can find $Q(\xi, \alpha^{**}), \alpha^{**} \in \Lambda(C)$, such that

$$d\left(\left(\inf_{\alpha \in \Lambda(C)} R_{if}(\alpha), R_{ifemp}(\alpha^{**})\right)_{\geq}\right) > \frac{\varepsilon}{2} \text{ and } R_{if}(\alpha^{**}) \geq \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha).$$

Then

$$d\left(\left(R_{if}(\alpha^{**}), R_{ifemp}(\alpha^{**})\right)_{\geq}\right) > \frac{\varepsilon}{2}.$$

Therefore

$$\begin{aligned} P(A_2) &\leq P\left\{d\left(\left(R_{if}(\alpha^{**}), R_{ifemp}(\alpha^{**})\right)_{\geq}\right) > \frac{\varepsilon}{2}\right\} \\ &\leq P\left\{\sup_{\alpha \in \Lambda} d\left(\left(R_{if}(\alpha), R_{ifemp}(\alpha)\right)_{\geq}\right) > \frac{\varepsilon}{2}\right\} \xrightarrow{l \rightarrow \infty} 0. \end{aligned}$$

Hence

$$P(A) \leq P(A_1) + P(A_2) \xrightarrow{l \rightarrow \infty} 0.$$

The theorem has been proved.

The samples become fuzzy random samples when $\nu_{\xi} = 1 - \mu_{\xi}$. Then we have

Corollary 3.1 [26]. Let the set of functions $Q(\xi, \alpha), \alpha \in \Lambda$, satisfies $X \leq R(\alpha) \leq Y$. Then the sufficient and necessary condition for the strict consistency of FERM principle is the convergence

$$\lim_{l \rightarrow \infty} P\left\{\sup_{\alpha \in \Lambda} D\left(\left(R_f(\alpha), R_{femp}(\alpha)\right)_{\geq}\right) > \varepsilon\right\} = 0$$

is valid.

The samples become random samples in the usual sense when the samples are not represented as fuzzy sets, and let d be the subtraction of real numbers. Then we have

Corollary 3.2 [47]. Assume that there exist the constants a and A such that for all functions in the set $Q(\xi, \alpha), \alpha \in \Lambda$, and for a given distribution function $F(\xi)$, the inequalities

$$a \leq R(\alpha) \leq A, \alpha \in \Lambda$$

hold true. Then the following two statements are equivalent:

1. For the given distribution function $F(\xi)$, the empirical risk minimization method is strictly consistent on the set of functions $Q(\xi, \alpha), \alpha \in \Lambda$.

311 2. For the given distribution function $F(\xi)$, the uniform one-sided convergence of the
 312 means to their mathematical expectation takes place over the set of functions $Q(\xi, \alpha), \alpha \in \Lambda$.

313

314 4. BOUNDS OF THE RATE OF UNIFORM CONVERGENCE OF INTUITIONISTIC 315 FUZZY RANDOM SAMPLES

316

317 In statistical learning theory, the important conclusions about the relationship between
 318 empirical risk and actual risk form the promotion of some boundary. They are essential to
 319 learning machine capacity analysis and the development of new learning algorithms. An
 320 important part of the rate of convergence of learning processes is the generalization bounds.
 321 In this section, consensus convergence of learning process based on intuitionistic fuzzy
 322 random samples is discussed. We consider the model where a set of intuitionistic fuzzy
 323 measurable functions $Q(\xi, \alpha), \alpha \in \Lambda$ contains a finite number N of elements
 324 $Q(\xi, \alpha_k), k = 1, 2, \dots, N$. At first, we introduce a basic theorem:

325 **Theorem 4.1**[25,26]. Let $\xi_j = \left\{ \left[(\xi_j)_r^-, (\xi_j)_r^+ \right] \mid r \in [0, 1] \right\}, j = 1, 2, \dots, n$ be a sequence of fuzzy

326 random variables, and $(\xi_j)_r^-, (\xi_j)_r^+$ have the limited variance and the same upper bounds M .

327 Then

$$328 \quad P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \xi_j, E \xi_j \right) \geq \varepsilon \right\} \leq \frac{2M}{n\varepsilon^2}.$$

329 **Theorem 4.2.** Let $\xi_j = \left\{ \langle x, \mu_{\xi_j}(x), \nu_{\xi_j}(x) \rangle \mid x \in U \right\}, j = 1, 2, \dots, n$ be a sequence of intuitionistic

330 fuzzy random variables, $(\mu_{\xi_j})_r^-, (\mu_{\xi_j})_r^+$ and $(\nu_{\xi_j})_r^-, (\nu_{\xi_j})_r^+$ have the limited variance and the

331 same upper bounds M_1 and M_2 , respectively. Then

$$332 \quad P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \xi_j, E \xi_j \right) \geq \varepsilon \right\} \leq \frac{8(M_1 + M_2)}{n\varepsilon^2}.$$

333 **Proof.** According to Theorem 4.1, we have

$$\begin{aligned}
& P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \xi_j, E \xi_j \right) \geq \varepsilon \right\} = P \left\{ \max \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right), d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \right\} \geq \varepsilon \right\} \\
& \leq P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) + d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \varepsilon \right\} \\
& \leq P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) + d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \varepsilon \right\} \\
& \leq P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) \geq \frac{\varepsilon}{2}, \text{ or }, d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} \\
& = P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} + P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} - \\
& P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \& d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} \\
& \leq P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} + P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} \\
& \leq \frac{8M_1}{n\varepsilon^2} + \frac{8M_2}{n\varepsilon^2} = \frac{8(M_1 + M_2)}{n\varepsilon^2}
\end{aligned}$$

334

335 **Theorem 4.3.** Suppose that $Q(\xi, \alpha_k), k=1, 2, \dots, N$ is a set of the functions, $E[Q(\xi, \alpha)]$
 336 exists and satisfies the conditions of Theorem 4.2. Then the following inequality

$$337 \quad d \left((R_{if}(\alpha_l), R_{ifemp}(\alpha_l))_{\geq} \right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}}$$

338 holds with probability of at least $1 - \eta$.

339 **Proof.**

$$\begin{aligned}
340 \quad & P \left\{ \sup_{1 \leq k \leq N} d \left((R_{if}(\alpha_k), R_{ifemp}(\alpha_k))_{\geq} \right) > \varepsilon \right\} \leq \sum_{k=1}^N P \left\{ d \left((R_{if}(\alpha_k), R_{ifemp}(\alpha_k))_{\geq} \right) > \varepsilon \right\} \\
& \leq N \cdot \frac{8(M_1 + M_2)}{l\varepsilon^2}
\end{aligned}$$

$$341 \quad \text{Let } N \cdot \frac{8(M_1 + M_2)}{l\varepsilon^2} = \eta, \text{ we conclude that } \varepsilon = \sqrt{\frac{8N(M_1 + M_2)}{l\eta}}.$$

342 Therefore

$$343 \quad P \left\{ d \left((R_{if}(\alpha_l), R_{ifemp}(\alpha_l))_{\geq} \right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}} \right\} \geq 1 - \eta.$$

344 **Theorem 4.4.** The inequality

$$345 \quad d \left((R_{if}(\alpha_l), R_{if}(\alpha_0))_{\geq} \right) \leq 2 \sqrt{\frac{2N(M_1 + M_2)}{l\eta}}$$

346 is satisfied with probability of at least $1 - 2\eta$.

347 **Proof.** Considering the properties of d and Theorem 4.3, we have

$$\begin{aligned}
& d\left(\left(R_f(\alpha_l), R_f(\alpha_0)\right)_\geq\right) \leq d\left(\left(R_f(\alpha_l), R_{femp}(\alpha_l)\right)_\geq\right) + d\left(R_{femp}(\alpha_l), R_f(\alpha_0)\right) \\
& \leq d\left(\left(R_f(\alpha_l), R_{femp}(\alpha_l)\right)_\geq\right) + d\left(\left(R_{femp}(\alpha_0), R_f(\alpha_0)\right)_\geq\right) \\
& \leq 2\sqrt{\frac{8N(M_1 + M_2)}{l\eta}}.
\end{aligned}$$

Let the following relationships hold

$$A: d\left(\left(R_f(\alpha_l), R_{femp}(\alpha_l)\right)_\geq\right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}};$$

$$B: d\left(\left(R_{femp}(\alpha_0), R_f(\alpha_0)\right)_\geq\right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}};$$

$$C: d\left(\left(R_f(\alpha_l), R_f(\alpha_0)\right)_\geq\right) \leq 2\sqrt{\frac{8N(M_1 + M_2)}{l\eta}},$$

and $P(A) \geq 1 - \eta; P(B) \geq 1 - \eta$.

We can conclude that C must hold true in virtue of A and B true, then $P(C) \geq P(AB)$.

Therefore

$$P(C) \geq P(AB) = 1 - P((AB)^c) \geq 1 - P(A^c) - P(B^c) \geq 1 - 2\eta.$$

The theorem is proved.

When the samples are the fuzzy random samples, let

$$R_f(\alpha_l) = d\left(\int Q(\xi, \alpha) dF(\xi), I_{\{0\}}\right) \text{ and } R_{femp}(\alpha_l) = d\left(\frac{1}{n} \sum_{k=1}^l Q(\xi_k, \alpha), I_{\{0\}}\right). \text{ Then we have}$$

Corollary 4.1[25,26]. Suppose that $Q(\xi, \alpha_k)$, $k = 1, 2, \dots, N$ is a set of functions

and $Q(\xi, \alpha_k)_r = \left[(Q(\xi, \alpha_k))_r^-, (Q(\xi, \alpha_k))_r^+ \right]$. If $(Q(\xi, \alpha_k))_r^-$, $(Q(\xi, \alpha_k))_r^+$ have the limited variance and the same upper bounds M , then

$$1) R_f(\alpha_l) - R_{femp}(\alpha_l) \leq \sqrt{\frac{2M}{l\eta}} \text{ holds true with probability of at least } 1 - \eta.$$

$$2) R_f(\alpha_l) - R_f(\alpha_0) \leq 2\sqrt{\frac{2M}{l\eta}} \text{ is satisfied with probability of at least } 1 - 2\eta.$$

Let d denote the subtraction of real numbers. We use Hoeffding's inequalities when the samples are the random samples in the usual sense. We have

Corollary 4.2[47]. If $\{Q(\xi, \alpha_k), \alpha_k \in \Lambda, k = 1, 2, \dots, N\}$ is bounded, i.e., $A \leq Q(\xi, \alpha_k) \leq B$, then

$$1) R(\alpha_l) - R_{emp}(\alpha_l) \leq (B - A) \sqrt{\frac{\ln N - \ln \eta}{2l}} \text{ is valid with probability of at least } 1 - \eta.$$

$$2) R(\alpha_l) - R(\alpha_0) \leq B \sqrt{\frac{\ln N - \ln \eta}{2l}} + (B - A) \sqrt{\frac{-\ln \eta}{2l}} \text{ holds true with probability of at least } 1 - 2\eta.$$

371

372 5. CONCLUSIONS

373

374 This paper discusses the intuitionistic fuzzy numbers and intuitionistic fuzzy random
 375 variables. We have proved the strong law of large numbers of intuitionistic fuzzy random
 376 variables and showed some related theorems and useful properties. Furthermore, based on

intuitionistic fuzzy random samples, we propose the principle of intuitionistic fuzzy empirical risk minimization of learning theory, prove the key theorem of learning theory based on intuitionistic fuzzy random samples, and discuss the bounds of the rate of uniform convergence of learning process. Altogether these findings have laid the foundation for further research in statistical learning theory involving intuitionistic fuzzy random samples. Further investigations might focus on such fundamental issues as intuitionistic fuzzy structural risk minimization and address the applied aspects such as support vector machines based on intuitionistic fuzzy random samples, etc.

ACKNOWLEDGEMENTS

The authors thank the anonymous referees for their valuable suggestions in improving this paper. This work is supported by the National Natural Science Foundation of China (Grant Nos. 61073121 and 61375075) and the Natural Science Foundation of Hebei Province of China (Grant Nos. F2012201020 and A2012201033).

COMPETING INTERESTS

AUTHORS' CONTRIBUTIONS

'Zhiming Zhang' designed the study, performed the statistical analysis, wrote the protocol, and wrote the overall draft of the manuscript. The author read and approved the final manuscript.

REFERENCES

1. K. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 20 (1986) 87–96.
2. K. Atanassov, More on intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 33 (1989) 37–45.
3. Atanassov, New operations defined over the intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 61 (1994) 137–142.
4. K. Atanassov, Remarks on the intuitionistic fuzzy sets – III, *Fuzzy Sets and Systems* 75 (1995) 401–402.
5. K. Atanassov, *Intuitionistic Fuzzy Sets: Theory and Applications*, Physica-Verlag, Heidelberg, 1999.
6. Zhaoqi Bian and Xuegong Zhang, *Pattern Recognition* (in Chinese), Tsinghua University Press, Beijing, 1999.
7. M. Brunato, R. Battiti, Statistical learning theory for location fingerprinting in wireless LANs. *Computer Networks* 47 (2005) 825–845.
8. P. Burillo, H. Bustince, Construction theorems for intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 84 (1996) 271–281.
9. H. Bustince, P. Burillo, Structures on intuitionistic fuzzy relations, *Fuzzy Sets and Systems* 78 (1996) 293–303.
10. J. L. Castro, L. D. Flores-Hidalgo, C. J. Mantas, J. M. Puche, Extraction of fuzzy rules from support vector machines, *Fuzzy Sets and Systems* 158 (2007) 2057–2077.
11. G. C. Cawley and N. L. C. Talbot, Improved sparse least-squares support vector machines, *Neurocomputing*, 48 (1-4) (2002) 1025–1031.
12. A. Celikyilmaz, I. B. Turksen, Fuzzy functions with support vector machines, *Information Sciences* 177 (2007) 5163–5177.
13. M. D. Cock, C. Cornelis, E.E. Kerre, Intuitionistic fuzzy relational images, *Studies in Computational Intelligence* 2 (2005) 129–145.

- 427 14. D. Coker, An introduction to intuitionistic fuzzy topological spaces, *Fuzzy Sets and*
 428 *Systems* 88 (1997) 81–89.
- 429 15. C. Cornelis, G. Deschrijver, E.E. Kerre, Implication in intuitionistic fuzzy and interval-
 430 valued fuzzy set theory: construction, classification, application, *International Journal of*
 431 *Approximate Reasoning* 35 (2004) 55–95.
- 432 16. S.K. De, R. Biswas, A.R. Roy, Some operations on intuitionistic fuzzy sets, *Fuzzy Sets*
 433 *and Systems* 114 (2000) 477–484.
- 434 17. S.K. De, R. Biswas, A.R. Roy, Some operations on intuitionistic fuzzy sets, *Fuzzy Sets*
 435 *and Systems* 114 (2000) 477–484.
- 436 18. M. Demirci, Axiomatic theory of intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 110
 437 (2000) 253–266.
- 438 19. G. Deschrijver, E.E. Kerre, On the position of intuitionistic fuzzy set theory in the
 439 framework of theories modelling imprecision, *Information Sciences* 177 (2007) 1860–1866.
- 440 20. T. Evgeniou, T. Poggio, M. Pontil, et al, Regularization and statistical learning theory for
 441 data analysis, *Computational Statistics & Data Analysis* 38 (2002) 421–432.
- 442 21. M. H. Ha, Y. C. Bai, P. Wang, et al, The key theorem and the bounds on the rate of
 443 uniform convergence of statistical learning theory on a credibility space, *Advances in Fuzzy*
 444 *Sets and Systems* 1 (2006) 143–172.
- 445 22. M. H. Ha, Z. F. Feng, S. J. Song, L. Q. Gao, The key theorem and the bounds on the
 446 rate of uniform convergence of statistical learning theory on quasi-probability spaces,
 447 *Chinese Journal of Computers* 31 (2008) 476–485.
- 448 23. M. H. Ha, Y. Li, J. Li, D. Z. Tian, The key theorem and the bounds on the rate of uniform
 449 convergence of learning theory on Sugeno measure space, *Science in China: Series F* 49
 450 (2006) 1–14.
- 451 24. M. H. Ha, W. Pedrycz, L. F. Zheng, The theoretical fundamentals of learning theory
 452 based on intuitionistic fuzzy random samples, *Fuzzy Sets and Systems* 178 (2009) 3240–
 453 3246.
- 454 25. Minghu Ha, Wenguang Tang and Yunchao Bai. The key theorem of learning theory
 455 about fuzzy samples, *Proceeding of the Second International Conferences on Machine*
 456 *Learning and Cybernetics*, Xian, 2–5 November 2003, 1213–1216. 3–7.
- 457 26. M. H. Ha, J. Tian, The theoretical foundations of statistical learning theory based on
 458 fuzzy number samples, *Information Sciences* 178 (2008) 3240–3246.
- 459 27. M. H. Ha, J. Tian, C. L. Zhou, The key theorem and bounds on the rate of convergence
 460 of statistical learning theory on the fuzzy samples, *Dynamics of Continuous, Discrete and*
 461 *Impulsive Systems, Series A: Mathematical Analysis* 14(S1) (2007) 72–76.
- 462 28. C. L. Huang, M.C. Chen, C.J. Wang, Credit scoring with a data mining approach based
 463 on support vector machines, *Expert System with Applications* 33 (2007) 847–856.
- 464 29. C. Hwang, D. H. Hong, K. H. Seok, Support vector interval regression machine for crisp
 465 input and output data, *Fuzzy Sets and Systems* 157 (2006) 1114–1125.
- 466 30. B. Jin, Y. C. Tang, Y. Q. Zhang, Support vector machines with genetic fuzzy feature
 467 transformation for biomedical data classification, *Information Sciences* 177 (2007) 476–489.
- 468 31. T. Kikuchi, S. Abe, Comparison between error correcting output codes and fuzzy support
 469 vector machines, *Pattern Recognition Letters* 26 (2005) 1937–1945.
- 470 32. Y. K. Kim, A strong law of large numbers for fuzzy random variables, *Fuzzy Sets and*
 471 *Systems* 111 (2000) 319–323.
- 472 33. D. Li, F. Shan, C. Cheng, On properties of four IFS operators, *Fuzzy Sets and Systems*
 473 154 (2005) 151–155.
- 474 34. Z. Z. Liang, P. F. Shi, Similarity measures on intuitionistic fuzzy sets, *Pattern Recognition*
 475 *Letters* 24 (2003) 2687–2693.

- 476 35. C. C. Lin, S. H. Chen, T. K. Truong and Y. Chang, Auto classification and categorization
477 based on wavelets and support vector machine, IEEE Transactions on Speech and Audio
478 Processing, 13 (5) (2005), 644-651.
- 479 36. C. F. Lin, S. D. Wang, Fuzzy support vector machines, IEEE Transactions on Neural
480 Networks 13 (2002) 464-471.
- 481 37. L. Lin, X. H. Yuan, Z. Q. Xia, Multicriteria fuzzy decision-making methods based on
482 intuitionistic fuzzy sets, Journal of Computer and System Sciences 73 (2007) 84-88.
- 483 38. B. D. Liu. Uncertainty Theory: An Introduction to its Axiomatic Foundations, Springer-
484 Verlag, Berlin, 2004.
- 485 39. Y. H. Liu, Y. T. Chen, Face recognition using total margin-based adaptive fuzzy support
486 vector machines, IEEE Transactions on Neural Networks 18 (2007) 178-192.
- 487 40. T. K. Mondal, S. K. Samanta, Topology of interval-valued intuitionistic fuzzy sets, Fuzzy
488 Sets and Systems 119 (2001) 483-494.
- 489 41. C. Su, C. H. Yang, Feature selection for the SVM: an application to hypertension
490 diagnosis, Expert Systems with Applications 34 (2008) 754-763.
- 491 42. D. Z. Tian, Z. M. Zhang, M. H. Ha, The key theorem of complex statistical learning theory,
492 Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis
493 14(S1) (2007) 46-50.
- 494 43. K. Tomonori and A. Shigeo, Comparison between error correcting output codes and
495 fuzzy support vector machines, Pattern Recognition Letters, 26 (12) (2005), 1937-1945.
- 496 44. C. F. Tsai, Training support vector machines based on stacked generalization for image
497 classification, Neurocomputing, 64 (2005), 497-503.
- 498 45. J. J. Tsong, Hybrid approach of selecting hyper-parameters of support vector machine
499 for regression, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics
500 36 (2006) 699-709.
- 501 46. V. N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York,
502 1995.
- 503 47. V. N. Vapnik, Statistical Learning Theory, A Wiley-Interscience Publication, New York,
504 1998.
- 505 48. V. N. Vapnik, An overview of statistical learning theory, IEEE Transactions on Neural
506 Networks 10 (1999) 988-999.
- 507 49. M. Vidyasagar, Statistical learning theory and randomized algorithms for control, IEEE
508 Control Systems Magazine 12 (1998) 69-85.
- 509 50. Y. Q. Wang, S. Y. Wang, K. K. Lai, A new fuzzy support vector machine to evaluate
510 credit risk, IEEE Transactions on Fuzzy Systems 13 (2005) 820-931.
- 511 51. Z. Y. Wang, J. K. George, Fuzzy Measure Theory, Plenum Press, New York, 1992.
- 512 52. H. Wechsler, Z. Duric, F. Y. Li, V. Cherkassky, Motion estimation using statistical
513 learning theory, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004)
514 466-478.
- 515 53. W. Y. N. Wing, A. Dorado, D. S. Yeung, W. Pedrycz and E. Izquierdo, Image
516 classification with the use of radial basis function neural networks and the minimization of
517 the localized generalization error, Pattern Recognition, 40 (2007), 19 -32.
- 518 54. Z.S. Xu, Intuitionistic preference relations and their application in group decision making,
519 Information Sciences 177 (2007) 2363-2379.
- 520 55. L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338-353.
- 521 56. Yiqiang Zhan and Dinggang Shen, Design efficient support vector machine for fast
522 classification, Pattern Recognition, 38 (1) (2005), 157-161.
- 523 57. C. Y. Zhang, H. Y. Fu, Similarity measures on three kinds of fuzzy sets, Pattern
524 Recognition Letters 27 (2006) 1307-1317.
- 525 58. Atanassov, K. Intuitionistic fuzzy sets, VII ITKR's Session, Sofia, June 1983 (Deposed
526 in Central Sci. -Techn. Library of Bulg. Acad. of Sci., 1697/84) (in Bulg.)

- 527 59. Atanassov, K. Remark on intuitionistic fuzzy numbers. Notes on Intuitionistic Fuzzy Sets,
528 Vol. 13, 2007, No. 3, 29-32.