A possibilistic analogue to Bayes estimation with fuzzy data and its application in machine learning

Mohsen Arefi · Reinhard Viertl · S. Mahmoud Taheri

Accepted: 3 March 2022 / Published online: 20 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

A Bayesian approach in a possibilistic context, when the available data for the underlying statistical model are fuzzy, is developed. The problem of point estimation with fuzzy data is studied in the possibilistic Bayesian approach introduced. For calculating the point estimation, we introduce a method without considering a loss function, and one considering a loss function. For the point estimation with a loss function, we first define a risk function based on a possibilistic posterior distribution, and then the unknown parameter is estimated based on such a risk function. Briefly, the present work extended the previous works in two directions: First the underlying model is assumed to be probabilistic rather than possibilistic, and second is that the problem of Bayes estimation is developed based on two cases of without and with considering loss function. Then, the applicability of the proposed approach to concept learning is investigated. Particularly, a naive possibility Bayes classifier is introduced and applied to some real-world concept learning problems.

Keywords

Lifetime data · Maximum possibilistic posterior estimator · Point estimation · Possibilistic Bayes approach · Possibilistic posterior distribution · Risk function

1 Introduction

Bayesian inference for parametric statistical models is based on two assumptions:

(1) The parameter of interest, \( \theta \), of the underlying model \( f(x|\theta) \) is of stochastic nature and has a probabilistic prior distribution, \( \pi(\theta) \).

(2) The available data related to the random variable \( X \) are precise.

The first item is the most important point in the Bayesian paradigm. But, considering \( \theta \) as a random variable with a probabilistic prior distribution is a matter of challenge between frequentist and Bayesian statisticians. On the other hand, concerning item 2 above, in many situations, the available data are vague (non-precise) rather than crisp (precise) (see Zimmermann 2000).

In real-world problems, we often face such situations, situations in which the interesting parameter has a possibilistic nature and/or the data available are fuzzy. For instance, consider a medical study about the rate of people affected by virus COVID-19 in a certain population. Suppose that, based on the experiences, this rate is determined to be “about 0.15” (expressed by a fuzzy number with certain spreads as ambiguity). Moreover, it is completely reasonable that some medical tests about presence or absence of the virus yield uncertain results; for example, “with high possibility” she/he has the virus, the possibility is “about 0.8” s to have virus, etc. In such a study, which is very usual in practical problems, we need to use a possibilistic version of the Bayes formula based on fuzzy information.

In this study, we propose a possibilistic version of the Bayesian approach, in which the underlying model is probabilistic, but the prior information about \( \theta \) is formulated as a...
possibility distribution (called possibilistic prior distribution) rather than a probabilistic one. In addition, we consider that the data available for the random variable $X$ are presented as fuzzy numbers rather than as crisp numbers. To this end, the likelihood function based on such a fuzzy-valued random sample is defined, and then, the extended likelihood function is combined with the possibilistic prior distribution, to obtain a possibilistic version of the posterior distribution. Such a possibilistic posterior distribution is based on a T-norm, so that it is flexible with respect to different situations. In this regard, the problem of point estimation without and with loss function is presented, too. Also, we present a possibilistic predictive distribution for predicting future fuzzy values of the underlying variable. In fact, we introduce a new Bayes classifier which is based on the possibility prior (rather than a probability one) and fuzzy data (rather than crisp ones). Therefore, the novelty of the current work compared to previous works is:

1) Considering the underlying model to be probabilistic rather than possibilistic.
2) Studying the problem of Bayes estimation based on two cases of without and with considering loss function, in a decision-theoretic framework.
3) Introducing the possibilistic predictive distribution function, in a probabilistic–possibilistic context.
4) Introducing the naive possibilistic Bayes classifier for crisp and fuzzy data.

The application of the proposed approach can be used in machine learning in which the Bayesian procedures present some methods to combine the prior knowledge and available data (see, e.g., Cui et al. 2013; He et al. 2014; Jiang et al. 2014; Subrahmanya and Shin 2013; Wang et al. 2018; You et al. 2019).

The topic of Bayesian inference in vague (non-precise) environments and/or in expert systems has been studied by some authors. Let us review some recent works in this topic. Lapointe and Bobée (2000) studied a possibilistic posterior distribution based on a possibilistic prior distribution and a possibilistic statistical model. Arefi and Taheri (2016) extended Lapointe and Bobée’s approach when the available data are fuzzy. They also studied the applicability of their approach in the field of concept learning. Taheri and Behboodian (2001, 2006) extended the Bayes approach to testing fuzzy hypotheses for both crisp (exact) and fuzzy (non-exact) data (also, see Torabi and Behboodian (2007). Some approaches of point estimations with fuzzy random samples are investigated by Akbari and Khanjari Sadegh (2012). Osoba et al. (2011) considered the problem of Bayesian inference using some certain fuzzy priors. Bacani and Barros (2017), Bardakhchyan (2017), Hareter and Viertl (2004), Mandal and Ranadive (2019), Osoba et al. (2012), Viertl (2011), Viertl and Hareter (2004), and Zhang and Chi (2008) investigated some aspects of the Bayes approaches in imprecise (fuzzy/vague) environments.

This paper is organized as follows: In Sect. 2, we recall some basic concepts of possibility theory. Two new concepts, the possibilistic prior distribution and possibilistic posterior distribution, are introduced and investigated in Sect. 3. In Sect. 4, we study the problem of parameter estimation, in the possibilistic Bayes paradigm, without considering a loss function. In Sect. 5, using a loss function, the posterior risk function is initially defined and then, the problem of parameter estimation is developed based on this function. The concept of possibilistic predictive distributions for future values of fuzzy data is introduced in Sect. 6. Some applications of the proposed model in machine learning are explained in Sect. 7. In Sect. 8, the proposed approach is compared with some other approaches. A brief conclusion is provided in Sect. 9.

2 Possibility measure and possibility function

The Bayesian approach in statistics is fundamentally based on considering the parameter of interest $\theta$ (related to the statistical model $f(x|\theta)$) as a random variable with a prior probabilistic distribution $\pi(\theta)$. However, in many problems, we have an imprecise (not necessarily stochastic) information on $\theta$. In these cases, it is reasonable to consider $\theta$ as a possibilistic variable with a vague (fuzzy) prior information. Below, we recall two basic definitions of “Possibility Theory,” which we will need in the present article. The reader is referred to Dubois and Prade (1988), Klir and Folger (1988), Krätschmer (2004), and Zadeh (1968) for more details.

Definition 1 A possibility measure $\Pi$ on a measurable space $(\Omega, \mathcal{B})$ is defined to be a function $\Pi: \mathcal{B} \to [0, 1]$, that satisfies the following axioms

i) $\Pi(A) \geq 0$, for all $A \in \mathcal{B}$,

ii) $\Pi(\emptyset) = 1$,

iii) $\Pi(\bigcup_{i=1}^{\infty} A_i) = \sup_i \Pi(A_i)$, (for every sequence $A_i \in \mathcal{B}$).

Also, $(\Omega, \mathcal{B}, \Pi)$ is said to be a possibility space.

Definition 2 Suppose that $(\Omega, \mathcal{B}, \Pi)$ is a possibility space. The function $\pi^*(\cdot): \Omega \to [0, 1]$ is a possibility function related to $\Pi$ if it satisfies

$$\Pi(A) = Poss(A) = \sup_{x \in A} \pi^*(x), \quad \forall A \in \mathcal{B}.$$ 

In the special case, $\pi^*(x) = \Pi(\{x\})$. 
Remark 1 The possibility measure and the possibility function are comparable to the probability measure \( P(\cdot) \) and the probability density function, respectively. Note that if \( f(\cdot) \) is a probability density function on \((\Omega, B)\), then we have

\[
P(A) = \int_A f(x)dx, \quad \forall A \in B.
\]

Remark 2 Note that the possibility measure \( \Pi(\cdot) \) and the probability measure \( P(\cdot) \) are set functions, but the possibility function \( \pi^*(\cdot) \) and the probability function \( f(\cdot) \) are real-valued functions defined on \( \Omega \).

3 Possibilistic posterior distribution with fuzzy data

In this section, we extend the concept of likelihood function to fuzzy data. Moreover, we define the possibilistic posterior distribution based on a possibilistic prior distribution, when the observations of the underlying model are fuzzy. In the following, we assume that \((\Omega, B)\) is a measurable space, in which \(\Omega\) is the sample space. Also, \((\Omega, B, P)\) is a probability space, where \(P\) is a probability measure on \((\Omega, B)\).

Definition 3 Let \((\Omega, B, P)\) be a probability space. Let \(\tilde{X} = (\tilde{X}_1, \tilde{X}_2, ..., \tilde{X}_n)\) be a fuzzy-valued random sample of size \(n\) of \(X\), associated with the probability density function (PDF) or a probability mass function \(f(\cdot|\theta)\), i.e., a sequence of fuzzy numbers as fuzzy realizations of the original random variable \(X\). Then, the likelihood function based on such a fuzzy-valued random sample is defined by

\[
\begin{align*}
 l(\theta; \tilde{X}) &= P(\tilde{X}|\theta) = \int X \int X \cdots \int X \prod_{i=1}^n [\tilde{X}_i(x_i) f(x_i|\theta)] \\
 &\quad \cdot dv(x_1)dv(x_2) \cdots dv(x_n),
\end{align*}
\]

where \(f(x|\theta)\) is the Radon–Nikodym derivative of \(P\) with respect to \(\nu\) (a \(\sigma\)-finite measure). The measure \(\nu\) usually is “counting measure” or “Lebesgue measure,” and \(X\) is the support of the random variable \(X\).

Remark 3 It should be mentioned that when the available data are crisp numbers \(x_1, x_2, ..., x_n\), then the above definition reduces to the ordinary definition of the likelihood function, i.e., \(l(\theta; x) = \prod_{i=1}^n f(x_i|\theta)\). Note that the above extension is according to Zadeh’s definition (Zadeh 1968) for the probability of fuzzy events.

Definition 4 Let \(f(x|\theta)\) be a statistical model with unknown parameter \(\theta \in \Theta\). Suppose that the information about \(\theta\) is formulated as a possibility function \(\pi^*(\theta)\). This possibility function is called possibilistic prior distribution for \(\theta\).

Definition 5 Consider the fuzzy-valued random sample \(\tilde{X} = (\tilde{X}_1, \tilde{X}_2, ..., \tilde{X}_n)\) with the likelihood function \(l(\theta; \tilde{X})\). Suppose that the parameter \(\theta\) has a possibilistic prior distribution \(\pi^*(\theta)\). The possibilistic posterior distribution, under T-norm \(T(\cdot, \cdot)\), is defined by

\[
\pi^*(\theta|\tilde{X}) = \frac{T(l(\theta; \tilde{X}), \pi^*(\theta))}{m(\tilde{X})},
\]

where \(m(\tilde{X}) = \sup_\theta \{T(l(\theta; \tilde{X}), \pi^*(\theta))\}\) is called the marginal function (for obtaining a normal posterior distribution, i.e., \(\exists \theta \text{ s.t. } \pi^*(\theta|\tilde{X}) = 1\)).

The flowchart of procedure for calculating the possibilistic posterior distribution is given in Fig. 1.

Remark 4 It is remarkable that in the above discussion, we have two kinds of uncertainty: the probabilistic one which is due to the random variable and the possibilistic one which is arisen from the possibilistic prior. The difficulty is to combine these two different kinds of uncertainties to update the information about the unknown parameter \(\theta\). Note that although there have been some attempts to define the conditional possibilities (e.g., Coletti and Vantaggi (1999), De Baets et al. (1999), Ferracuti and Vantaggi (2006), Hisdal (1978), Kramosil (1998), and Nguyen (1978)), to the best knowledge of the authors, there has not been any work in the problem of updating a possibility information using probabilistic fuzzy data.

Remark 5 The above definition is, in some sense, consistent to some definitions for conditional possibility. First, note that the marginal function \(m(\tilde{X})\) is analogous to a corresponding relation for probabilities in probability theory for which the operators summation and product are used instead of sup and \(T(\cdot, \cdot)\), respectively (see Eq. (33), in Nguyen (1978), Eq. (2.8) in Hisdal (1978), and Eq. (6) in Kramosil (1998)). A second item is related to use a T-norm in the numerator in Eq. (2).

Example 1 The data in Table 1 (centers of the fuzzy numbers) show the lifetimes (in 1000 km) of front disk brake pads on a randomly selected set of 40 cars (same model) that were monitored by a dealer network (see, Lawless, 2003, p. 337).
Suppose that the lifetime of the front disk brake pad has an exponential distribution with density function
\[ f(t|\theta) = \frac{1}{\theta} e^{-\frac{t}{\theta}}, \quad t > 0, \theta > 0, \]
where \( \theta \) is the mean lifetime of the front disk brake pad. An expert believes that the value of the variable \( \theta \) lies in the interval \([40,50]\) with a possibility of one. Moreover, he/she believes it is possible that \( \theta \) is smaller than 40, but never below 30, and bigger than 50, but never above 60. We use a trapezoidal fuzzy number to model the possibilistic prior distribution based on the expert opinion as follows:

\[
\pi^*(\theta) = \begin{cases} 
\frac{\theta - 30}{10} & 30 \leq \theta \leq 40, \\
1 & 40 < \theta \leq 50, \\
\frac{60 - \theta}{10} & 50 < \theta \leq 60, \\
0 & \text{otherwise}.
\end{cases}
\]

In practice, measuring the lifetime of a disk may not yield an exact result. A disk may work perfectly over a certain period but be braking for some time and finally be unusable at a certain time. So, such data may be reported as imprecise quantities. Assume that the lifetimes of front disk brake pads are reported as fuzzy numbers in Table 1. In fact, imprecision is formulated by fuzzy numbers \( \tilde{X}_i = (x_i, s_i) \), with \( s_i = 0.05 x_i, i = 1, 2, \ldots, 40 \), as follows:

\[
\tilde{X}_i(t) = \begin{cases} 
1 - \frac{t - x_i}{s_i} & x_i \leq t \leq x_i + s_i, \\
0 & \text{otherwise}.
\end{cases}
\]

The likelihood function based on such fuzzy data is calculated as
\[
l(\theta; \tilde{X}) = \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^n \left[ \tilde{X}_i(t_i) \frac{1}{\theta} e^{-\frac{t_i}{\theta}} \right] dt_1 \cdots dt_n
\]
\[
= e^{-\sum_{i=1}^n \frac{t_i}{\theta}} \cdot \prod_{i=1}^n \left[ 1 + \frac{\theta (e^{-s_i}/\theta - 1)}{s_i} \right]
\]
\[
= e^{-\frac{1051}{\theta}} \cdot \prod_{i=1}^{40} \left[ 1 + \frac{\theta (e^{-s_i}/\theta - 1)}{s_i} \right].
\]

Hence, the possibilistic posterior distribution based on the product T-norm, \( T(a, b) = a \cdot b \), is obtained as follows (see Fig. 2)

\[
\pi^*(\theta | \tilde{X}) = \begin{cases} 
\frac{\theta - 30}{10 - m(\tilde{X})} \cdot e^{-\frac{2051}{\theta}} \cdot \prod_{i=1}^{40} \left[ 1 + \frac{\theta (e^{-s_i}/\theta - 1)}{s_i} \right] & 30 \leq \theta \leq 40, \\
\frac{1}{m(\tilde{X})} \cdot e^{-\frac{2051}{\theta}} \cdot \prod_{i=1}^{40} \left[ 1 + \frac{\theta (e^{-s_i}/\theta - 1)}{s_i} \right] & 40 < \theta \leq 50, \\
\frac{60 - \theta}{10 - m(\tilde{X})} \cdot e^{-\frac{2051}{\theta}} \cdot \prod_{i=1}^{40} \left[ 1 + \frac{\theta (e^{-s_i}/\theta - 1)}{s_i} \right] & 50 < \theta \leq 60,
\end{cases}
\]

where
\[
m(\tilde{X}) = \sup_{0 < \theta} \left\{ l(\theta; \tilde{X}) \cdot \pi^*(\theta) \right\} = 1.6012 \times 10^{-83}.
\]

Example 2 Consider a study for estimating the proportion \( \theta \) of a kind of tree in a forest, which is infected with a plague. We take a sample of 3 trees and examine each tree separately for the presence of plague. Suppose that we have no precise mechanisms for an exact distinction between the presence or absence of plague, but we can identify the information with a fuzzy set on \( X = \{0, 1\} \).

The usual model for this problem starts from the Bernoulli experiment \( X \) associated with the presence of plague (\( Bin(1, \theta), 0 \leq \theta \leq 1 \)). Suppose that, based on some prior information, we know that \( \tilde{\theta} \) is approximately 0.2. In this case, a suitable possibilistic distribution for modeling such prior information is the following fuzzy number (see Fig. 3)
A possibilistic analogue to Bayes estimation with fuzzy data and its application... 5501

\[ \pi^*(\theta) = \begin{cases} 
25\theta^2 & 0 \leq \theta \leq 0.2, \\
\frac{25}{16}(1 - \theta)^2 & 0.2 < \theta \leq 1.
\end{cases} \]

Suppose that, based on a random sample of size \( n = 3 \), we observe the following fuzzy data

\[ X_1 = \left\{ \frac{0.2}{0}, \frac{0.9}{1} \right\}, 
X_2 = \left\{ \frac{0.1}{0}, \frac{0.8}{1} \right\}, 
X_3 = \left\{ \frac{0.4}{0}, \frac{0.6}{1} \right\}. \]

The likelihood function based on these fuzzy data is (\( t = x_1 + x_2 + x_3 \))

\[ l(\theta; X) = \frac{1}{x_1=0} \sum_{x_2=0}^{x_3=0} \left[ (0.9)^{x_1} \cdot (0.2)^{1-x_1} \cdot (0.8)^{x_2} \cdot (0.1)^{1-x_2} \cdot (0.6)^{x_3} \cdot (0.4)^{1-x_3} \right] \theta^t \cdot (1 - \theta)^3 - t = 0.008 (1 - \theta)^3 + 0.112 \theta (1 - \theta)^2 + 0.438 \theta^2 (1 - \theta) + 0.432 \theta^3. \]

Based on the product T-norm, the possibilistic posterior distribution would be (see Fig. 4)

\[ \pi^*(\theta|X) = \begin{cases} 
\frac{1}{m(\theta,X)} \left[ \frac{1}{2} \theta^2 (1 - \theta)^3 + \frac{14}{5} \theta^3 (1 - \theta)^2 \\
+ \frac{210}{20} \theta^4 (1 - \theta) + \frac{54}{5} \theta^5 \right] & 0 \leq \theta \leq 0.2, \\
\frac{1}{m(\theta,X)} \left[ \frac{1}{20} \theta^2 (1 - \theta)^3 + \frac{7}{20} \theta (1 - \theta)^4 \\
+ \frac{210}{350} \theta^2 (1 - \theta)^3 + \frac{27}{30} \theta^3 (1 - \theta)^2 \right] & 0.2 < \theta \leq 1,
\end{cases} \]

which is maximized in \( \theta = 0.4340 \). Also, the marginal function for the observed data is

\[ m(\bar{X}) = \sup_{0 \leq \theta \leq 1} (l(\theta; \bar{X}) \cdot \pi^*(\theta|\bar{X}) = (l(\theta = 0.4340; \bar{X}) \cdot \pi^*(\theta = 0.4340) = 0.0496. \]

4 Point estimation without loss functions

**Definition 6** Consider a possibilistic posterior distribution based on an observed fuzzy random sample. Then, \( \hat{\theta} = d(\bar{X}) \), as a function on the fuzzy-valued random sample \( \bar{X} = (\bar{X}_1, \bar{X}_2, ..., \bar{X}_n) \) to \( R \), is called the maximum possibilistic posterior estimation of \( \theta \) (\( MPPE(\theta) \)) if

\[ \pi^*(\hat{\theta}|\bar{X}) \geq \pi^*(\theta|\bar{X}), \quad \forall \theta \in \Theta. \]

The above definition is similar to the definition of the maximum Bayesian likelihood estimator, defined as the posterior mode, in the probabilistic approach (see, Robert (2001), p. 166).

In spite of Definition 6, we can use defuzzification methods (see, i.e., Ross (1995)) to obtain an estimation for \( \theta \) based on the possibilistic posterior distribution.

**Definition 7** Let \( DFF(\cdot) \) be a defuzzification operator. The point estimation \( \hat{\theta} \) based on the possibilistic posterior distri-
bution \( \pi^*(\theta | \tilde{X}) \) is defined as
\[
\hat{\theta} = D \cdot F (\pi^*(\theta | \tilde{X})).
\] (4)

i) Consider the “Center of Gravity” defuzzification method. Based on this method, the point estimation \( \hat{\theta} = d_{COG}(\tilde{X}) \) is obtained as
\[
\hat{\theta} = d_{COG}(\tilde{X}) = \frac{\int_{\Theta} \theta \cdot \pi^*(\theta | \tilde{X}) d\theta}{\int_{\Theta} \pi^*(\theta | \tilde{X}) d\theta} = \frac{\int_{\Theta} \theta \cdot T(l(\theta; \tilde{X}), \pi^*(\theta)) d\theta}{\int_{\Theta} T(l(\theta; \tilde{X}), \pi^*(\theta)) d\theta}.
\]

ii) Consider the “Center of Area” defuzzification method. Based on this method, the point estimation \( \hat{\theta} = d_{COA}(\tilde{X}) \) is obtained as
\[
\int_{-\infty}^{\infty} \pi^*(\theta | \tilde{X}) d\theta = \int_{d_{COA}(\tilde{X})} \pi^*(\theta | \tilde{X}) d\theta,
\]
or
\[
\int_{-\infty}^{d_{COA}(\tilde{X})} T(l(\theta; \tilde{X}), \pi^*(\theta)) d\theta = \int_{d_{COA}(\tilde{X})} T(l(\theta; \tilde{X}), \pi^*(\theta)) d\theta.
\]

iii) Consider the “Supremum of Center” defuzzification method. Based on this method, the point estimation \( \hat{\theta} = d_{SOC}(\tilde{X}) \) is obtained as
\[
\hat{\theta} = d_{SOC}(\tilde{X}) = \sup_{\theta \in \Theta} \{ \theta \cdot \pi^*(\theta | \tilde{X}) \} = \sup_{\theta \in \Theta} \{ \theta \cdot T(l(\theta; \tilde{X}), \pi^*(\theta)) \}.
\]

**Example 3** Consider the possibilistic posterior distribution in Example 1. The point estimations of \( \hat{\theta} \), based on the above different methods, are calculated as follows:

- **The maximum possibilistic posterior estimation:**
  \( \hat{\theta}_1 = MPP\cdot (\theta) = 50. \)

- **The center of gravity estimation:**
  \( \hat{\theta}_2 = d_{COG}(\tilde{X}) = \frac{\int_{0}^{\infty} \theta \cdot \pi^*(\theta | \tilde{X}) d\theta}{\int_{0}^{\infty} \pi^*(\theta | \tilde{X}) d\theta} = \frac{576.2920}{11.8144} = 48.7788. \)

- **The center of area estimation:**
  \( \int_{0}^{d_{COA}} \pi^*(\theta | \tilde{X}) d\theta = \int_{0}^{40} \pi^*(\theta | \tilde{X}) d\theta + \int_{d_{COA}}^{\infty} \pi^*(\theta | \tilde{X}) d\theta = 5.9072, \)
  hence
  \( \hat{\theta}_3 = d_{COA}(\tilde{X}) = 49.0261. \)

5 Point estimation based on a loss functions

In a decision-theoretic analysis, the quality of a decision function (here, a point estimation) is quantified by the risk function of the decision function. In this section, we first define a risk function based on the possibilistic posterior distribution \( \pi^*(\theta | \tilde{X}) \), and then, the point estimation (decision function) \( d(\tilde{X}) \) is obtained based on this risk function.

Let \( \Theta \) be the parameter space. Any function \( L(\theta, d) : \Theta \times D \rightarrow \mathbb{R} \) is called a loss function, where \( D \) is the space of possible decisions (here, the space of all point estimations of \( \theta \)).
**Definition 8** The posterior risk function with fuzzy data $\tilde{X}$ for the estimation (decision function) $d(\tilde{X})$ under the probability density function (or probability mass function) $f(x|\theta)$ and with the possibilistic prior distribution $\pi^*(\theta)$, based on a loss function $L(\theta, d)$, is defined as

$$r(\pi^*(\theta|\tilde{X}), d) := \sup_{\theta \in \Theta} \left\{ L(\theta, d) \cdot \pi^*(\theta|\tilde{X}) \right\}. \quad (5)$$

**Definition 9** The estimation $d_{PB}$ based on the loss function $L(\theta, d)$ and the possibilistic posterior distribution based on the fuzzy data $\pi^*(\theta|\tilde{X})$ are called a possibilistic Bayes estimation if

$$r(\pi^*(\theta|\tilde{X}), d_{PB}) := \min_{d \in D} r(\pi^*(\theta|\tilde{X}), d), \quad (6)$$

where $D$ is the set of all estimations for $\theta$.

**Example 5** Consider the possibilistic posterior distribution in Example 1. The point estimation based on a loss function is calculated as follows:

(i) Based on the quadratic loss function $L(\theta, d) = (\theta - d)^2$, the posterior risk function is as in Fig. 5. Here, the possibilistic Bayes estimation of $\theta$, for which the posterior risk function is minimized, is $d_{PB} = 48.9355$.

(ii) Based on the loss function $L(\theta, d) = |\theta - d|$, the posterior risk function in terms of $d$ is as in Fig. 5. Here, the possibilistic Bayes estimation of $\theta$, for which the posterior risk function is minimized, is $d_{PB} = 49.1694$.

**Example 6** Consider the possibilistic posterior distribution in Example 2. The point estimation based on a loss function is calculated as follows:

i) Based on the quadratic loss function $L(\theta, d) = (\theta - d)^2$, the posterior risk function in terms of $d$ for the random sample of size $n = 3$ is shown in Fig. 6. Here, the possibilistic Bayes estimation of $\theta$, for which the posterior risk function is minimized, is $d_{PB} = 0.4491$.

ii) Based on the loss function $L(\theta, d) = |\theta - d|$, the posterior risk function for the random sample of size $n = 3$ is shown in Figure 6. Here, the possibilistic Bayes estimation of $\theta$, for which the posterior risk function is minimized, is $d_{PB} = 0.4315$.

### 6 Possibilistic predictive distribution

The future values of fuzzy data $\tilde{Y}$ are provided by a possibility function, called “possibilistic predictive distribution” (ppd).

**Definition 10** Let $\pi^*(\theta|\tilde{X})$ be the possibilistic posterior distribution based on the fuzzy data $\tilde{X} = (\tilde{X}_1, ..., \tilde{X}_n)$. Also, suppose that the random variable $Y$ has the pdf $f(\cdot|\theta)$. The possibilistic predictive distribution $g(\cdot|\tilde{X})$ is defined as

$$g(y|\tilde{X}) = \sup_{\theta \in \Theta} \left\{ T\left( \frac{f(y|\theta)}{f(y|\tilde{X})}, \pi^*(\theta|\tilde{X}) \right) \right\}, \quad (7)$$

where $f(y|\theta) = \sup_{\theta \in \Theta} \{ f(y|\theta) \}$.

**Remark 6** In the Bayesian approach, the predictive distribution for future values is introduced as

$$g(y|\tilde{X}) = \int \pi(\theta) \cdot \pi(\theta|\tilde{X}) \, d\theta,$$

where $\pi(\theta|\tilde{X})$ is the probabilistic posterior distribution based on the precise data $\tilde{X} = (x_1, ..., x_n)$ and $f(y|\theta)$ is the probability density function for the random variable $Y$ (see, Robert, 2001, p. 22). We list some differences of the possibilistic predictive distribution $g(y|\tilde{X})$ and the predictive distribution $g(y|\tilde{X})$ as follows:

1. The predictive distribution $g(y|\tilde{X})$ is based on precise data, but the possibilistic predictive distribution $g(y|\tilde{X})$ is based on fuzzy data.

2. The predictive distribution $g(y|\tilde{X})$ is a probability distribution, but the possibilistic predictive distribution $g(y|\tilde{X})$ is a possibility distribution.

3. The possibilistic predictive distribution $g(y|\tilde{X})$ is flexible under different $T$-norms.

**Example 7** Consider the possibilistic posterior distribution $\pi^*(\theta|\tilde{X})$ based on the product T-norm, in Example 1. The possibilistic predictive distribution $g(\cdot|\tilde{X})$ based on the product T-norm, $T(a, b) = a \cdot b$, is obtained as follows (see Fig. 7)

$$g(t|\tilde{X}) = \sup_{\theta > 0} \left\{ T\left( \frac{f(t|\theta)}{f(t|\tilde{X})}, \pi^*(\theta|\tilde{X}) \right) \right\} = \sup_{\theta > 0} \left\{ \frac{T}{\theta} \cdot \pi^*(\theta|\tilde{X}) \right\}$$

The possibilistic predictive distribution can be expressed by “approximately 50.” Table 2 shows some values of the possibilistic predictive distribution. The new front disk brake pad works 50 (in 1000 km) with possibility 1, and it works less/more than 50 with less possibility. (For example, with possibility 90%, the lifetime of the disk is equal to 30.)

**Example 8** Consider Example 2. The possibilistic predictive distribution based on some T-norms is obtained as follows:

i) Based on the product T-norm, $T(a, b) = a \cdot b$, the possibilistic posterior distribution is obtained as follows (see Example 2)
Fig. 5 The posterior risk function in Example 5

![Image of Figure 5](image1.png)

Fig. 6 The posterior risk functions in Example 6

![Image of Figure 6](image2.png)

Table 2 Some values of possibilistic predictive distribution in Example 7

| $t$  | 10  | 20  | 30  | 40  | 50  |
|------|-----|-----|-----|-----|-----|
| ppd  |    |    |    |    |    |
| 0.4451 | 0.7288 | 0.8951 | 0.9771 | 1   |

| $t$  | 60  | 70  | 80  | 90  | 100 |
|------|-----|-----|-----|-----|-----|
| ppd  |    |    |    |    |    |
| 0.9825 | 0.9384 | 0.8781 | 0.8088 | 0.7358 |

Fig. 7 The possibilistic predictive distribution in Example 7

![Image of Figure 7](image3.png)

Hence, the possibilistic predictive distribution is

$$
\pi^*(\theta | \tilde{X}) = \begin{cases}
\frac{1}{m(\tilde{X})} \left[ \frac{1}{3} \theta^2 (1-\theta)^3 + \frac{14}{3} \theta^3 (1-\theta)^2 \\
+ \frac{213}{320} \theta^4 (1-\theta)^3 + \frac{27}{320} \theta^3 (1-\theta)^4 \\
+ \frac{213}{320} \theta^2 (1-\theta)^3 + \frac{37}{320} \theta^3 (1-\theta)^4 \\
+ \frac{213}{320} \theta^2 (1-\theta)^3 + \frac{27}{320} \theta^3 (1-\theta)^4 \\
\right]
0 \leq \theta \leq 0.2, \\
\frac{25}{320} \theta^2 (1-\theta)^2 + \frac{54}{320} \theta^3 (1-\theta)^2 \\
0.2 \leq \theta \leq 1.
\end{cases}
$$

Fig. 8 Possibilistic posterior distribution based on the minimum T-norm in Example 8

![Image of Figure 8](image4.png)

\[ g(y|\tilde{X}) = \sup_{\theta \in \Theta} \left\{ T \left( \frac{f(y|\theta)}{f(y|\tilde{X})}, \pi^*(\theta|\tilde{X}) \right) \right\} \]

ii) Based on the minimum T-norm, \( T(a, b) = \min(a, b) \), the possibilistic posterior distribution is obtained as follows (see Fig. 8)

\[ \pi^*(\theta|\tilde{X}) = \frac{\min \left\{ l(\theta; \tilde{X}), \pi^*(\theta) \right\}}{m(\tilde{X})} \]

\[ = \begin{cases}
\frac{1}{16 \times 0.1923} \left[ 0 \leq \theta \leq 0.0198, \\
0.0198 < \theta \leq 0.6492, \\
0.6492 < \theta \leq 1.
\end{cases}
\]

Hence, the possibilistic predictive distribution is given by (see Fig. 9)

\[ g(y|\tilde{X}) = \sup_{\theta \in \Theta} \left\{ T \left( \frac{f(y|\theta)}{f(y|\tilde{X})}, \pi^*(\theta|\tilde{X}) \right) \right\} \]
7 Application in machine learning

A common problem in machine learning is the concept learning problem, for which the Bayes classifier is a well-known method, (see, e.g., He et al. (2014), Mitchell (1997), and Wang et al. (2014)). Any system that classifies new instances according to the following equation is called a Bayes optimal classifier or Bayes optimal learner (see Mitchell (1997))

\[ v_{MAP} = \arg\max_{\theta_j \in V} \pi(\theta_j | a_1, a_2, ..., a_n), \]

where \( V = \{\theta_1, ..., \theta_k\} \), \( a_1, a_2, ..., a_n \) are the attribute values, and \( \pi(\theta_j | .) \) is the probabilistic posterior function. Using Bayes rule, we can rewrite the above expression in the following way,

\[ v_{MAP} = \arg\max_{\theta_j \in V} \frac{P(a_1, a_2, ..., a_n | \theta_j) \pi(\theta_j)}{P(a_1, a_2, ..., a_n)} = \arg\max_{\theta_j \in V} P(a_1, a_2, ..., a_n | \theta_j) \pi(\theta_j). \]

Particularly, the naive Bayes classifier assumes that the attribute values are conditionally independent given the target value. In this case, the so-called naive Bayes classifier is defined as

\[ v_{NB} = \arg\max_{\theta_j} \prod_{i=1}^{n} P(a_i | \theta_j). \]

Now, based on the results in Sect. 3, we investigate and develop the Bayes classifier in the possibility environment. Particularly, we investigate naive possibilistic Bayes classifier both for crisp information and for fuzzy information.

7.1 Naive possibilistic Bayes classifier for crisp data

The possibilistic Bayesian approach to classify the new crisp instance is to assign the most target value of maximum a possibilistic posteriori (\( v_{MAP} \)) given the crisp attribute values \( \{a_1, a_2, ..., a_n\} \) as follows,

\[ v_{MAP} = \arg\max_{\theta_j \in V} \pi^*(\theta_j | a_1, a_2, ..., a_n) = \arg\max_{\theta_j \in V} \frac{T(P(a_1, a_2, ..., a_n | \theta_j), \pi(\theta_j))}{\sup_{\theta_j \in V} T(P(a_1, a_2, ..., a_n | \theta_j), \pi(\theta_j))}. \]

where \( V = \{\theta_1, ..., \theta_k\} \) and \( \pi(\theta_j) \) is the possibilistic prior distribution.

If the attribute values \( a_1, ..., a_n \) are conditionally independent given the target value, then the joint conditional probability is obtained as follows,

\[ P(a_1, a_2, ..., a_n | \theta_j) = \prod_{i=1}^{n} P(a_i | \theta_j). \]

The following example is a possibilistic version of the example in Mitchell (1997), p. 157, about a medical diagnosis.

Example 9 In this example, we investigate a learning task in a possibilistic context rather than probabilistic one. Consider a target attribute Medical Diagnosis with two values: (1) The patient has a particular form of cancer, and (2) the patient does not. The available data are from a laboratory test with two possible outcomes: \( \oplus \) (the result of test is positive) and \( \ominus \) (the result of test is negative.) Formally, the set of target values is \( V = \{\theta_1, \theta_2\} \), where \( \theta_1 = cancer \) and \( \theta_2 = nocancer \). Based on the prior information, a patient has cancer and nocancer with the possibilities \( \pi(\theta_1) = 0.10 \) and \( \pi(\theta_2) = 0.95 \), respectively. Here, the amount \( \pi(\theta_1) \) is the consistency of the symptoms of cancer in this patient (note that these values are not related to the randomness or relative frequency, which are usual in the probabilistic context).

The test returns a correct positive result in only 98% of the cases in which the disease is actually present and a correct negative result in only 97% of the cases in which the disease is not present. In other cases, the test returns the opposite result. The conditional probabilities are summarized as

\[ P(\oplus | \theta_1) = 0.98, \quad P(\ominus | \theta_1) = 0.02, \]

\[ P(\oplus | \theta_2) = 0.03, \quad P(\ominus | \theta_2) = 0.97. \]

Now, suppose that the laboratory test of a new patient returns a positive result. Our task is to diagnose the target value (the patient has cancer or not) based on the possibilistic posterior distribution.
7.2 Naive possibilistic Bayes classifier for fuzzy data

In this subsection, we extend the method in the previous subsection to the case when the observed data are fuzzy rather than crisp. We use the possibilistic posterior distribution to classify the new fuzzy instance. Generally, suppose that the new fuzzy instance $\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n$ is available and we want to classify it in the target space $V = \{\theta_1, ..., \theta_k\}$. We first calculate the possibilistic posterior distribution $\pi^*(\theta_j|\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n)$ based on Definition 5, and then, the most target value of the possibilistic posteriori ($v_{MAPP}$) is obtained as follows,

$$v_{MAPP} = \arg\max_{\theta_j \in V} \pi^*(\theta_j|\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n) = \arg\max_{\theta_j \in V} T(\theta|\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n, \pi^*(\theta)).$$

Here, the possibilistic posterior distribution is obtained as follows,

$$\pi^*(\theta_j|\oplus) = \frac{T(P(\oplus|\theta_j), \pi(\theta_j))}{\max[T(P(\oplus|\theta_1), \pi(\theta_1)), T(P(\oplus|\theta_2), \pi(\theta_2))]},$$

where $\theta_1 = \theta_1$, and $\theta_2$.

The results based on different T-norms are given in Table 3.

Based on this information, $v_{MAPP} = \theta_1$. Hence, if the result of test in a new patient is positive, then our best estimate is that he/she has cancer.

### Table 3

| $T(a, b)$ | $\theta = \theta_1$ | $\theta = \theta_2$ |
|-----------|---------------------|---------------------|
| $\min(a, b)$ | 1 | 0.30 |
| $a, b$ | 1 | 0.29 |
| $\max(0, a + b - 1)$ | 1 | 0 |
| $\frac{a, b}{a + b - 1}$ | 1 | 0.28 |

7.2 Naive possibilistic Bayes classifier for fuzzy data

In this subsection, we extend the method in the previous subsection to the case when the observed data are fuzzy rather than crisp. We use the possibilistic posterior distribution to classify the new fuzzy instance. Generally, suppose that the new fuzzy instance $\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n$ is available and we want to classify it in the target space $V = \{\theta_1, ..., \theta_k\}$. We first calculate the possibilistic posterior distribution $\pi^*(\theta_j|\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n)$ based on Definition 5, and then, the most target value of the possibilistic posteriori ($v_{MAPP}$) is obtained as follows,

$$v_{MAPP} = \arg\max_{\theta_j \in V} \pi^*(\theta_j|\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n) = \arg\max_{\theta_j \in V} T(\theta|\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n, \pi^*(\theta)).$$

The following example is a possibilistic version of the example in Mitchell (1997), p. 178, for the target concept “PlayTennis” (see also Arefi and Taheri (2016)).

### Example 10

Consider the learning task represented by the training examples of Table 4. Here, the target attribute PlayTennis, which can have values yes or no for different Saturday mornings, is to be predicted based on other attributes of the morning in question. Formally, the set of target values is $V = \{\theta_1, \theta_2\}$, where $\theta_1 = Yes$ and $\theta_2 = No$. Based on the prior information, the player selects $\theta_1$ and $\theta_2$ with the possibilities $\pi(\theta_1) = 0.91$ and $\pi(\theta_2) = 0.15$, respectively. Let each day be described by the following attribute values,

$\theta_1: \text{Outlook} = \{\text{Sunny, Overcast, Rain}\}$,
$\theta_2: \text{Temperature} = \{\text{Cool, Mild, Hot}\}$,
$\theta_3: \text{Humidity} = \{\text{Normal, High}\}$,
$\theta_4: \text{Wind} = \{\text{Weak, Strong}\}$.

Based on the training examples given in Table 4, the conditional probabilities are obtained as in Table 5.

Suppose that we observe the following fuzzy information for a new instance,
Based on Definition 3, the likelihood function under $\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4$ is calculated as

$$l(\theta; (\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4)) = P(\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4|\theta)$$

$$= \prod_{i=1}^{4} P(\tilde{a}_i|\theta)$$

$$= \prod_{i=1}^{4} \left[ \sum_{x_i} \tilde{a}_i(x_i) P(x_i|\theta) \right]$$

$$= \left[ \frac{63}{50} \times \frac{62}{50} \times \frac{57}{50} \times \frac{60}{50} \right] \theta = Yes,$$

$$\left[ \frac{57}{50} \times \frac{57}{50} \times \frac{57}{50} \times \frac{57}{50} \right] \theta = No,$$

$$= \left[ \frac{2036}{2036} \right] \theta = Yes,$$

$$\left[ \frac{1131}{1131} \right] \theta = No.$$
Viertl (2011) proposed a Bayesian inference method in the situation of fuzzy prior information and fuzzy data. He combined fuzzy sample $x_1^*, \ldots, x_n^*$ with the related membership functions $\xi_1(\cdot), \ldots, \xi_n(\cdot)$. First, he combined the membership functions of the fuzzy data to a membership function $\zeta(\cdot, \ldots, \cdot)$ of the combined fuzzy sample $\chi^*$. Then, he used the extension principle on the likelihood function $l(\cdot | x_1, \ldots, x_n)$, the characterizing (membership) function $\eta(\cdot)$ of the fuzzy value $l(\theta | x_1^*, \ldots, x_n^*)$, based on fuzzy data $x_1^*, \ldots, x_n^*$, is given by

$$
\eta(y) = \begin{cases}
\sup(\zeta(X)) : x \in M_n^\pi, l(\theta | x) = y & \text{if } l^{-1}(\{y\}) \neq \emptyset
\end{cases},
$$

for all $y \in R$.

In order to obtain the generalized Bayes’ theorem, he used the $\delta$-level functions $\pi_\delta(\cdot)$ and $\pi_\delta(\cdot)$ of $\pi^*(\cdot)$ and $\pi^*(\cdot)$ and $\pi_\delta^*(\cdot)$ of $l(\cdot | x^*)$. Hence, a fuzzy posterior density $\pi^*(\cdot | x^*)$ is calculated by its $\delta$-level functions $\pi_\delta(\cdot | x^*)$ and $\pi_\delta(\cdot | x^*)$ as follows

$$
\begin{align*}
\pi_\delta(\theta | x^*) &= \int_0^\delta \frac{\pi_\delta(\theta) \pi_\delta^*(\theta | x^*)}{\int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*) + \int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*)} d\theta \\
\pi_\delta^*(\theta | x^*) &= \frac{\int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*) + \int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*)}{\int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*) + \int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*)} d\theta
\end{align*}
$$

$\forall \theta \in \Theta, \delta \in (0, 1]$.

Both the methods proposed in this article and in Viertl’s work are based on the likelihood function, but our method has the following advantages:

1) In our method, the parameter $\theta$ has a possibilistic nature, but it has a probabilistic nature in Viertl’s. Hence, the posterior distribution based on our method is possibilistic–probabilistic in nature, whereas it is fuzzy probabilistic in the Viertl method.

2) The posterior distribution proposed in the present paper is based on T-norms. Hence, the posterior distribution is flexible using different T-norms.

3) In our method, we investigate the case in which a loss function has been considered in estimating the unknown parameter.

8.1 Comparison with the Gil et al.’s work

Gil et al. (1985) studied the point estimation problem with fuzzy information. They defined a probabilistic posterior distribution as follows

$$
\pi(\theta | \chi) = \frac{P(\chi | \theta) \cdot \pi(\theta)}{\int_\Theta P(\chi | \theta) \cdot \pi(\theta) d\theta},
$$

where $P(\chi | \theta)$ is defined in Definition 3, and $\pi(\theta)$ is a probabilistic prior distribution. Also, the posterior risk function based on the loss function $L(\theta, d)$ for the estimated parameter $\theta$ is defined as follows

$$
\mathcal{R}(\pi, d) = E_{\theta|\chi}(L(\theta, d(\chi))) = \int_\Theta L(\theta, d(\chi)) d\pi(\theta | \chi).
$$

In contrast with the Gil et al.’s method, it should be noted that:

1) In our method, the parameter $\theta$ has a possibilistic nature, but it has a probabilistic nature in Gil et al.’s. Hence, in our method, the posterior distribution will also be of a possibilistic–probabilistic nature, whereas it will be of a probabilistic nature in Gil et al.’s.

2) We define the posterior distribution based on a T-norm. Hence, the posterior distribution is flexible under the different T-norms.

8.2 Comparison with Viertl’s work

Viertl (2011) proposed a Bayesian inference method in the situation of fuzzy prior information and fuzzy data. He supposed that a fuzzy sample $x_1^*, \ldots, x_n^*$ consists of fuzzy numbers with related membership functions $\xi_1(\cdot), \ldots, \xi_n(\cdot)$. First, he combined the membership functions of the fuzzy data to obtain a membership function $\zeta(\cdot, \ldots, \cdot)$ of the combined fuzzy sample $\chi^*$. Then, based on the extension principle on the likelihood function $l(\cdot | x_1, \ldots, x_n)$, the characterizing (membership) function $\eta(\cdot)$ of the fuzzy value $l(\theta | x_1^*, \ldots, x_n^*)$, based on fuzzy data $x_1^*, \ldots, x_n^*$, is given by

$$
\eta(y) = \begin{cases}
\sup(\zeta(X)) : x \in M_n^\pi, l(\theta | x) = y & \text{if } l^{-1}(\{y\}) \neq \emptyset
\end{cases},
$$

for all $y \in R$.

In order to obtain the generalized Bayes’ theorem, he used the $\delta$-level functions $\pi_\delta(\cdot)$ and $\pi_\delta(\cdot)$ of $\pi^*(\cdot)$ and $\pi^*(\cdot)$ and $\pi_\delta^*(\cdot)$ of $l(\cdot | x^*)$. Hence, a fuzzy posterior density $\pi^*(\cdot | x^*)$ is calculated by its $\delta$-level functions $\pi_\delta(\cdot | x^*)$ and $\pi_\delta(\cdot | x^*)$ as follows

$$
\begin{align*}
\pi_\delta(\theta | x^*) &= \int_0^\delta \frac{\pi_\delta(\theta) \pi_\delta^*(\theta | x^*)}{\int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*) + \int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*)} d\theta \\
\pi_\delta^*(\theta | x^*) &= \frac{\int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*) + \int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*)}{\int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*) + \int_0^\delta \pi_\delta(\theta) \pi_\delta^*(\theta | x^*)} d\theta
\end{align*}
$$

$\forall \theta \in \Theta, \delta \in (0, 1]$.

Both the methods proposed in this article and in Viertl’s work are based on the likelihood function, but our method has the following advantages:

1) In our method, the parameter $\theta$ has a possibilistic nature, but it has a probabilistic nature in Viertl’s. Hence, the posterior distribution based on our method is possibilistic–probabilistic in nature, whereas it is fuzzy probabilistic in the Viertl method.

2) The posterior distribution proposed in the present paper is based on T-norms. Hence, the posterior distribution is flexible using different T-norms.

3) In our method, we investigate the case in which a loss function has been considered in estimating the unknown parameter.
A possibilistic analogue to Bayes estimation with fuzzy data and its application...

where $T(\cdot, \cdot)$ is a T-norm. Then, based on a possibilistic prior distribution $\pi(\theta)$, the possibilistic posterior distribution is defined by

$$
\pi(\theta|\tilde{X}) = \psi_T(\pi(\tilde{X}), \pi(\tilde{X}, \theta)),
$$

where $\pi(\tilde{X}, \theta) = T(\pi(\theta), \pi(\tilde{X}|\theta))$, $\pi(\tilde{X}) = \sup_{\theta \in \Theta} (\pi(\tilde{X}, \theta))$, and $\psi_T(a, c) = \sup_{x \in [0, 1]; T(a, c) \leq c}$ is a residuated implication operator (see also Lapointe and Bobée (2000)). Some cases are the following:

1. The approaches introduced in Lapointe and Bobée (2000) (with crisp data) and Arefi and Taheri (2016) (with fuzzy data) are full possibilistic nature, i.e., the model of data and the prior distribution have possibilistic distributions, but our method is a combined version of probability and possibility, i.e., the model of data is probabilistic and the prior distribution is possibilistic.

2. In this paper, we investigate the case in which a loss function has been considered in estimating the unknown parameter.

9 Conclusions

In the Bayesian method for estimation of an unknown parameter, it is often difficult to assuming the prior information in stochastic terms. In addition, sometimes the observed data are non-precise (fuzzy) rather than precise (crisp). In this paper, by introducing the concept of likelihood function for fuzzy data of a probabilistic model, a possibilistic approach was described for dealing with such situations. The present approach employed the possibility distribution for modeling the prior information. Then, based on the possibilistic posterior distribution, some methods were proposed to estimate the unknown parameter of interest with/without a loss function. The proposed approach was applied to a well-known problem in machine learning called the concept learning problem. Particularly, the naive possibility Bayes classifier is introduced and applied to some real-world concept learning problems.

In summary, we can express the contributions of the paper as follows:

1. Introducing a new approach to combination of probabilistic model and possibilistic information to get a possibilistic posterior distribution function.
2. Considering the observed data of the underlying model as fuzzy rather than crisp.
3. Employing a decision-theoretic approach in which the problem of estimation was studied based on two cases of without and with considering loss function. In addition, the maximum possibilistic posterior estimation (as a duality of the concept of maximum Bayesian likelihood estimation) is introduced.
4. Introducing the possibilistic predictive distribution function, based on probabilistic model, fuzzy observations, and possibilistic prior information.
5. Investigating some applications of the proposed Bayesian approach in the problem of concept learning, by introducing the naive possibilistic Bayes classifier for crisp data and fuzzy data.

It should be mentioned that, according to the proposed approach, choosing (or constructing) a suitable prior distribution is an important issue (see, e.g., Hill and Spall (1994), Chen et al. (2000), and Dubois (2006)), which is a potential subject of future research.

Acknowledgements The authors wish to express their thanks to the referee for valuable comments. The third author would like to acknowledge the financial support of University of Tehran for this research under grant number 30005/1/05.

Funding The authors have not disclosed any funding.

Data availability The data used in the presented work are included within the paper.

Declarations

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical approval This article does not contain any studies with human participants or animals.

References

Akbari MG, Khanjari Sadegh M (2012) Estimators based on fuzzy random variables and their mathematical properties. Iran J Fuzzy Syst 9:79–95
Arefi M, Taheri SM (2016) Possibilistic Bayesian inference based on fuzzy data. Int J Mach Learn Cybern 7:753–763
Bacani F, Barros LC (2017) Application of prediction models using fuzzy sets: a Bayesian inspired approach. Fuzzy Sets Syst 319:104–116
De Baets B, Tsiporkova E, Mesiar R (1999) Conditioning in possibility theory with strict order norms. Fuzzy Sets Syst 106:221–229
Bardakchyan VG (2017) Fuzzy Bayesian inferences. Phys Math Sci 51:8–12
Chen YH, Wang WJ, Chiu CH (2000) New estimation method for the membership values in fuzzy sets. Fuzzy Sets Syst 112:521–525
Coletti G, Vantaggi B (2009) T-conditional possibilities: coherence and inference. Fuzzy Sets Syst 160:306–324
Cui Y, Guo R, Guo D (2013) Probabilistic DEAR models. Int J Mach Learn Cybern 4:373–389
Dubois D (2006) Possibility theory and statistical reasoning. Comput Statist Data Anal 51:47–69
Dubois D, Prade H (1988) Possibility theory. Plenum Press, New York
Ferracuti L, Vantaggi B (2006) Independence and conditional possibility for strictly monotone triangular norms. Int J Intel Syst 21:299–323
Gil MA, Corral N, Gil P (1985) The fuzzy decision problem: an approach to the point estimation problem with fuzzy information. Eur J Oper Res 22:26–34
Hareter D, Viertl R (2004) Fuzzy information and Bayesian statistics. Soft methodology and random information systems, advances in soft computing, Springer, Berlin, Heidelberg 26, 392–398
He YL, Wang R, Kwong S, Wang XZ (2014) Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis. Inf Sci 259:252–268
Hill SD, Spall JC (1994) Sensitivity of a Bayesian analysis to the prior distribution. IEEE Trans Syst Man Cybern 24:216–221
Hisdal E (1978) Conditional possibilities, independence and noninteraction. Fuzzy Sets Syst 1:283–297
Jiang L, Cai Z, Wang D, Zhang H (2014) Bayesian citation-KNN with distance weighting. Int J Mach Learn Cybern 5:193–199
Klir GJ, Folger T (1988) Fuzzy Sets. Uncertainty and Information, Prentice-Hall, Englewood Cliffs, NJ
Kramosil I (1998) Alternative definitions of conditional possibilistic measures. Kybernetika 34:137–147
Krätschmer V (2004) Probability theory in fuzzy sample spaces. Metrika 60:167–189
Lapointe S, Bobée B (2000) Revision of possibility distributions: a Bayesian inference pattern. Fuzzy Sets Syst 116:119–140
Lawless JF (2003) Statistical models and methods for lifetime data, 2nd edn. Wiley, New Jersey
Mandal P, Ranadive AS (2019) Multi-granulation fuzzy probabilistic rough sets and their corresponding three-way decisions over two universes. Iran J Fuzzy Syst 16:61–76
Mitchell TM (1997) Machine learning. McGraw Hill, New York
Nguyen HT (1978) On conditional possibility distributions. Fuzzy Sets Syst 1:299–309
Osoba O, Mitaim S, Kosko B (2011) Bayesian inference with adaptive fuzzy priors and likelihoods. IEEE Trans Syst Man Cybern Part B Cybern 41:1183–1197
Osoba O, Mitaim S, Kosko B (2012) Triply fuzzy function approximation for hierarchical Bayesian inference. Fuzzy Optim Decis Making 11:241–268
Robert CP (2001) The Bayesian choice, 2nd edn. Springer, Heidelberg
Ross TJ (1995) Fuzzy logic with engineering applications. Mcgraw-Hill, New York
Subrahmanya N, Shin YC (2013) A variational Bayesian framework for group feature selection. Int J Mach Learn Cybern 4:609–619
Taheri SM, Behboodian J (2001) A Bayesian approach to fuzzy hypotheses testing. Fuzzy Sets Syst 123:39–48
Taheri SM, Behboodian J (2006) On Bayesian approach to testing fuzzy hypotheses with fuzzy data. Italian J Pure Appl Math 19:139–154
Torabi H, Behboodian J (2007) Likelihood ratio tests for fuzzy hypotheses testing. Statist Papers 48:509–522
Viertl R (2011) Statistical methods for fuzzy data. Wiley, Chichester
Viertl R, Hareter D (2004) Generalized Bayes’ theorem for non-precise a-priori distribution. Metrika 59:263–273
Wang XZ, He YL, Wang DD (2014) Non-naive Bayesian classifiers for classification problems with continuous attributes. IEEE Trans Cybern 44:21–39
Wang J, Zhang L, Cao JJ, Han D (2018) NBWELM: naive Bayesian based weighted extreme learning machine. Int J Mach Learn Cybern 9:21–35
You Y, Li J, Shen L (2019) An effective Bayesian network parameters learning algorithm for autonomous mission decision-making under scarce data. Int J Mach Learn Cybern 10:549–561
Zadeh LA (1968) Probability measures of fuzzy events. J Math Anal Appl 23:421–427
Zhang Y, Chi ZX (2008) A Fuzzy support vector classifier based on Bayesian optimization. Fuzzy Optim Decis Making 7:75–86
Zimmermann HJ (2000) An application-oriented view of modeling uncertainty. Eur J Oper Res 122:190–198

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.