Possibilistic Bayesian Models

Mohsen Arefi[1], Reinhard Viertl[2], and S. Mahmoud Taheri[3]
[1]Department of Statistics, Faculty of Mathematical Sciences and Statistics, University of Birjand, Birjand, Iran, Arefi@birjand.ac.ir
[2]Institute of Statistics and Probability Theory, Vienna University of Technology, Vienna, Austria, r.viertl@tuwien.ac.at
[3]Department of Mathematical Sciences, Isfahan University of Technology, Isfahan 84156-83111, Iran, Taheri@cc.iut.ac.ir

Abstract

The problem of modeling and analyzing fuzzy data is investigated in a possibilistic context, based on a Bayesian approach. Specially, we focus on the problem of point estimation when the available data of the underlying statistical model are fuzzy rather than crisp. To do this, first we extend the concept of likelihood function to fuzzy data. Then, to obtain the point estimation, we develop a method without considering a loss function and one considering a loss function based on a possibilistic posterior distribution. A few numerical examples are presented to explain the applicability of the proposed approach.

Keywords: Bayes approach, Likelihood function, Point estimation, Possibilistic posterior distribution, Possibility measure

1 Introduction and preliminary concepts

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

1) The parameter of interest, θ, 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 θ as a random variable with a probabilistic prior distribution is a matter of challenge between classical and Bayesian statisticians. On the other hand, concerning Item 2 above, in many situations, the available data are vague rather than crisp (see Zimmermann, 2000; and Tsoukiàs, 2008).

In this study, we propose a possibilistic version of the Bayesian approach, in which the prior information about θ is formulated as a possibility distribution rather than a probabilistic one. In addition, we consider that the data available for the random variable X are presented as fuzzy sets rather than as crisp numbers. Using this approach, we try to remove the above two limitations.

It is remarkable that, the Bayesian approach in statistics is fundamentally based on considering the parameter of interest θ (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 θ. In these cases, it is reasonable to consider θ 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 (2006) for more details.

Definition 1 *A possibility measure* Π *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) $\Pi(X) = 1$,

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* Π *if it satisfies*

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

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* $(\Omega, \mathcal{B})$*, then we have*

$$P(A) = \int_A f(x)dx, \qquad \forall\ A \in \mathcal{B}.$$

This paper is organized as follows: Two new concepts, the possibilistic prior distribution and possibilistic posterior distribution, are introduced and investigated in Section 2. In Section 3, we study the problem of parameter estimation, in the possibilistic Bayes paradigm, without considering a loss function. In Section 4, using a loss function, the posterior risk function is initially defined and then, the problem of parameter estimation is developed based on this function. A brief conclusion is provided in Section 5.

2 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 $(\Omega, \mathcal{B})$ is a measurable space, in which Ω is the sample space. Also, $(\Omega, \mathcal{B}, P)$ is a probability space, where P is a probability measure on $(\Omega, \mathcal{B})$.

Definition 3 *Let* $(\Omega, \mathcal{B}, P)$ *be a probability space. Suppose that* $\underline{\widetilde{X}} = (\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_n)$ *is 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*

$$l(\theta; \underline{\widetilde{X}}) = \int_{\chi} \cdots \int_{\chi} \prod_{i=1}^{n} \widetilde{X}_i(x_i) f(x_i|\theta) d\nu(x_i), \tag{1}$$

where, $f(x|\theta)$ *is the Radon Nikodym derivative of* P *with respect to* ν *(a* σ*-finite measure) and* χ *is the support of the random variable* X*.*

Remark 2 *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.* $l(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i|\theta)$*. 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 the unknown parameter $\theta \in \Theta$. Suppose that the information about θ is formulated as a possibility function $\pi^*(\theta)$. This possibility function is called the possibilistic prior distribution for θ.*

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

$$\pi^*(\theta|\underline{\widetilde{X}}) = \frac{T(l(\theta; \underline{\widetilde{X}}), \pi^*(\theta))}{m(\underline{\widetilde{X}})}, \qquad (2)$$

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

Remark 3 *The above definition is, in some sense, consistent to some definitions for conditional possibility. First, note that the marginal function $m(\underline{\widetilde{X}})$ is analogous to a corresponding relation for probabilities in probability theory for which the operators* sup *and $T(.,.)$ are used instead of summation and product, respectively, (see Equation (33) in Nguyen (1978) and Equation (6) in Kramosil (1998)). A second item is related to use a T-norm in the numerator in Equation (2). In this regard, see the discussion in Section 3 in De Baets et al. (1999) and Sections 3 and 4 in Coletti and Vantaggi (2009).*

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 the density function*

$$f(t|\theta) = \frac{1}{\theta} e^{-\frac{t}{\theta}}, \qquad t > 0, \qquad \theta > 0,$$

where θ is the mean lifetime of the front disk brake pad. An expert believes that the value of the variable θ lies in the interval [40,50] with a possibility of one. Moreover, he/she believes it is possible that θ 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 (see Figure 1)

$$\pi^*(\theta) = \begin{cases} \frac{\theta-30}{10} & 30 \le \theta \le 40, \\ 1 & 40 < \theta \le 50, \\ \frac{60-\theta}{10} & 50 < \theta \le 60, \\ 0 & \textit{otherwise}. \end{cases}$$

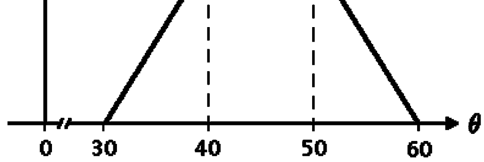

Figure 1. The possibilistic prior distribution in Example 1.

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* $\widetilde{X}_i = (x_i, s_i)_R$*, with* $s_i = 0.05x_i$*,* $i = 1, 2, \ldots, 40$*, as follows*

$$\widetilde{X}_i(t) = \begin{cases} 1 - \frac{t-x_i}{s_i} & x_i \le t \le x_i + s_i, \\ 0 & otherwise. \end{cases}$$

Table 1: Lifetimes of front disk brake pads of cars in Example 1

$No.$	$\widetilde{X}_i$	$No.$	$\widetilde{X}_i$	$No.$	$\widetilde{X}_i$	$No.$	$\widetilde{X}_i$
1	$(86.2, 4.3)_R$	11	$(36.7, 1.8)_R$	21	$(61.5, 3.1)_R$	31	$(45.9, 2.3)_R$
2	$(38.4, 1.9)_R$	12	$(22.6, 1.1)_R$	22	$(42.7, 2.1)_R$	32	$(50.6, 2.5)_R$
3	$(45.5, 2.3)_R$	13	$(81.7, 4.1)_R$	23	$(46.9, 2.3)_R$	33	$(59.0, 3.0)_R$
4	$(22.7, 1.1)_R$	14	$(102.5, 5.1)_R$	24	$(33.9, 1.7)_R$	34	$(62.4, 3.1)_R$
5	$(48.8, 2.4)_R$	15	$(28.4, 1.4)_R$	25	$(54.2, 2.7)_R$	35	$(34.4, 1.7)_R$
6	$(42.8, 2.1)_R$	16	$(31.7, 1.6)_R$	26	$(81.3, 4.1)_R$	36	$(50.2, 2.5)_R$
7	$(73.1, 3.7)_R$	17	$(52.1, 2.6)_R$	27	$(51.6, 2.6)_R$	37	$(50.7, 2.5)_R$
8	$(59.8, 3.0)_R$	18	$(56.4, 2.8)_R$	28	$(38.8, 1.9)_R$	38	$(64.5, 3.2)_R$
9	$(45.1, 2.3)_R$	19	$(42.2, 2.1)_R$	29	$(53.6, 2.7)_R$	39	$(33.8, 1.7)_R$
10	$(41.0, 2.1)_R$	20	$(40.0, 2.0)_R$	30	$(80.6, 4.0)_R$	40	$(56.7, 2.8)_R$

The likelihood function based on such fuzzy data is calculated as

$$\begin{aligned} l(\theta; \underline{\widetilde{X}}) &= \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^n \left[\widetilde{X}_i(t_i) \ \frac{1}{\theta} \ e^{-\frac{t_i}{\theta}} \right] \ dt_1 \cdots dt_n \\ &= e^{-\frac{2051}{\theta}} \cdot \prod_{i=1}^{40} \left[1 + \frac{\theta(e^{-s_i/\theta} - 1)}{s_i} \right]. \end{aligned}$$

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

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

where,

$$m(\underline{\widetilde{X}}) = \sup_{0 < \theta} \left\{ l(\theta; \underline{\widetilde{X}}) \cdot \pi^*(\theta) \right\} = l(\theta = 50; \underline{\widetilde{X}}) \cdot \pi^*(50) = 1.6012 \times 10^{-83}.$$

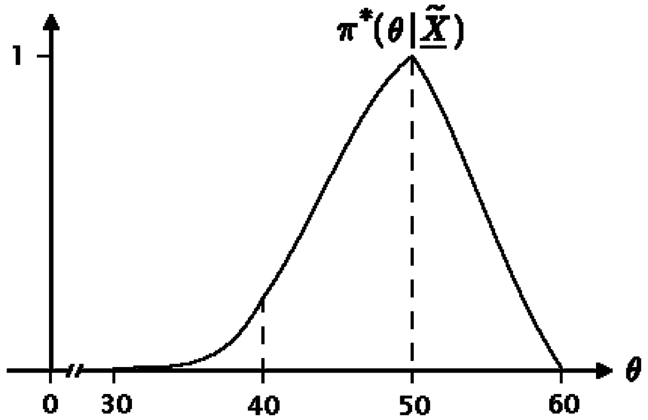

Figure 2. The possibilistic posterior distribution in Example 1.

3 Point estimation without loss functions

Definition 6 *Consider a possibilistic posterior distribution based on an observed fuzzy random sample. Then,* $\hat{\theta} = d(\underline{\widetilde{X}})$ *is called the maximum possibilistic posterior estimation of* θ *(*$MPPE(\theta)$*) if*

$$\pi^*(\hat{\theta}|\underline{\widetilde{X}}) \geq \pi^*(\theta|\underline{\widetilde{X}}), \qquad \forall \theta \in \Theta.$$

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).

Example 2 *Consider the possibilistic posterior distribution in Example 1. The maximum possibilistic posterior estimation of* θ *is calculated as* $\hat{\theta}_1 = MPPE(\theta) = 50$.

4 Point estimation based on a loss functions

In this section, we first define a risk function based on the possibilistic posterior distribution $\pi^*(\theta|\underline{\widetilde{X}})$, and then the point estimation (decision function) $d(\underline{\widetilde{X}})$ is obtained based on this risk function. Let Θ be the parameter space. Any function $L(\theta, d) : \Theta \times D \to \Re$ is called a loss function, where D is the space of possible decisions (here, the space of all estimations of θ).

Definition 7 *The posterior risk function with the fuzzy data* $\underline{\widetilde{X}}$ *for the estimation (decision function)* $d(\underline{\widetilde{X}})$ *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*

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

Definition 8 *The estimation* d^{PB} *based on the loss function* $L(\theta, d)$ *and the possibilistic posterior distribution for fuzzy data* $\pi^*(\theta|\underline{\widetilde{X}})$ *is called a possibilistic Bayes estimation if*

$$r(\pi^*(\theta|\underline{\widetilde{X}}), d^{PB}) := \min_{d \in D} r(\pi^*(\theta|\underline{\widetilde{X}}), d),$$

where D *is the set of all estimations for* θ.

Example 3 *Consider the possibilistic posterior distribution in Example 1. Then,*

i) *Based on the quadratic loss function* $L(\theta, d) = (\theta - d)^2$*, the posterior risk function is as in Figure 3. Here, the possibilistic Bayes estimation of* θ*, for which the posterior risk function is minimized, is* $d^{PB} = 48.9355\cdot$

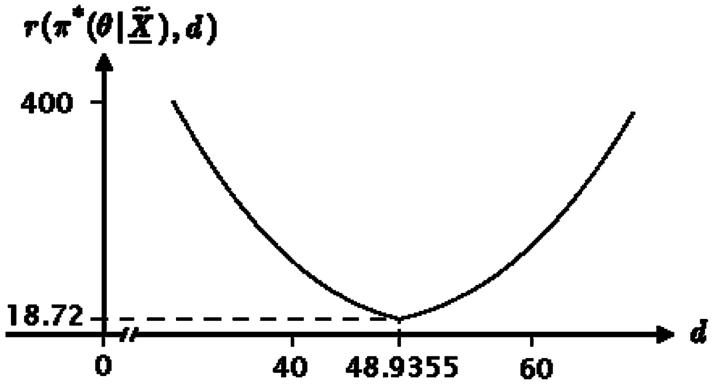

Figure 3. The posterior risk function based on $L(\theta, d) = (\theta - d)^2$ *in Example 3.*

ii) *Based on the loss function* $L(\theta, d) = |\theta - d|$, *the posterior risk function in terms of* d *is as in Figure 4. Here, the possibilistic Bayes estimation of* θ, *for which the posterior risk function is minimized, is* $d^{PB} = 49.1694$.

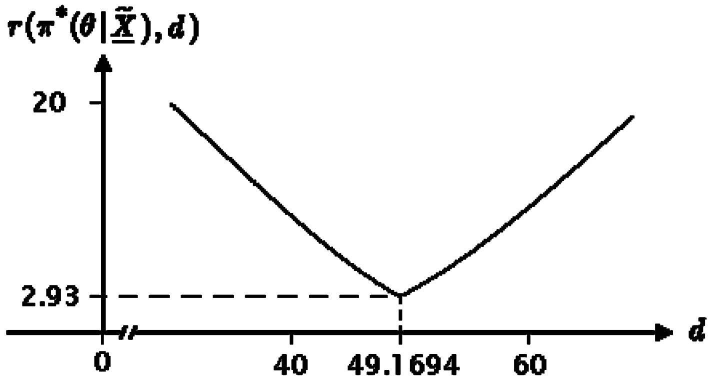

Figure 4. The posterior risk function based on $L(\theta, d) = |\theta - d|$ *in Example 3.*

5 Conclusion

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). This paper, by introducing the concept of likelihood function for fuzzy data, described a possibilistic approach for dealing with such situations. The presented approach uses the possibility distribution for modelling the prior information. Then, based on the possibilistic posterior distribution, we proposed some methods to estimate the unknown parameter of interest with/without a loss function.

References

De Baets, B., Tsiporkova, E., and Mesiar, R. (1999) "Conditioning in possibility theory with strict order norms," *Fuzzy Sets and Systems,* 106, 221-229.

Coletti, G. and Vantaggi, B. (2009) "T-conditional possibilities: Coherence and inference," *Fuzzy Sets and Systems,* 160, 306-324.

Dubois, D. (2006) "Possibility theory and statistical reasoning," *Computational Statistics & Data Analysis,* 51, 47-69.

Kramosil, I. (1998) "Alternative definitions of conditional possibilistic measures," *Kybernetika,* 34, 137-147.

Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data,* Sec. Ed., J. Wiley.

Nguyen, H.T. (1978) "On conditional possibility distributions," *Fuzzy Sets and Systems,* 1, 299-309.

Robert, C.P. (2001) *The Bayesian Choice,* (Second Edition), Springer-Verlag, Heidelberg.

Tsoukiàs, A. (2008) "From decision theory to decision aiding methodology," *European Journal of Operational Research,* 187, 138-161.

Zadeh, L.A. (1968) "Probability measures of fuzzy events," *Journal of Mathematical Analysis and Applications,* 23, 421-427.

Zimmermann, H.J. (2000) "An application-oriented view of modeling uncertainty," *European Journal of Operational Research,* 122, 190-198.