Journal of Uncertain Systems
Vol.2, No.3, pp.187-191, 2008
Online at: www.jus.org.uk

Foundations of Fuzzy Bayesian Inference

Reinhard Viertl*
Department of Statistics and Probability Theory, Vienna University of Technology
Wiedner Hauptstraße 8-10, 1040 Wien, Austria

Received 1 January 2008; Accepted 15 February 2008

Abstract

In applications of Bayesian inference frequently data and a-priori information are not precise numbers and not standard probability distributions on the parameter space, but more or less fuzzy. Therefore suitable descriptions of data are so-called non-precise numbers which are more general than fuzzy numbers, and so-called fuzzy probability distributions on the parameter space. Based on this it is necessary to generalize Bayes' theorem to this situation. This is possible and described in the paper.

Keywords: Bayes'theorem, fuzzy a-posteriori distributions, fuzzy a-priori distributions, fuzzy data, fuzzy predictive distributions

1 Introduction

The most up-to-date mathematical description of non-precise data is by special fuzzy subsets of the real line $\mathbb{R}$, called non-precise numbers.

A *non-precise number* $x^\star$ is a fuzzy subset of $\mathbb{R}$ whose membership function $\xi(\cdot)$ obeys the following conditions:

(1) $\forall \quad \delta \in (0;1]$, the δ-cut $C_\delta\left[\xi(\cdot)\right] := \{x \in \mathbb{R} \colon \xi(x) \geq \delta\}$ is a finite union of compact intervals $[a_{\delta,j}; b_{\delta,j}]$, i.e.

$$C_\delta\left[\xi(\cdot)\right] = \bigcup_{j=1}^{k_\delta} \left[a_{\delta,j}; b_{\delta,j}\right].$$

(2) $C_1[\xi(\cdot)] \neq \emptyset$.

A function $\xi(\cdot)$ fulfilling conditions (1) and (2) is called *characterizing function* of the non-precise number $x^\star$.

Remark 1: Special non-precise numbers are so-called *fuzzy numbers*. For them the δ-cuts are all non-empty compact intervals, i.e. $C_\delta\left[\xi(\cdot)\right] = \left[a_\delta; b_\delta\right]$ for $\forall \quad \delta \in (0;1]$.

For the generalization of Bayesian inference based on fuzzy samples $x_1^\star, \cdots, x_n^\star$ of a stochastic quantity X, for example a lifetime, it is necessary to consider the *sample space* M_X^n, where M_X is the set of possible values of X, called *observation space* M_X. In order to do that the concept of fuzzy vectors is necessary.

A *n-dimensional fuzzy vector* is a fuzzy subset of the n-dimensional Euclidean space $\mathbb{R}^n$ whose membership function $\zeta(\cdot, \cdots, \cdot)$ obeys the following conditions:

($\underline{1}$) $\forall \quad \delta \in (0;1]$, the δ-cut $C_\delta\left[\zeta(\cdot, \cdots, \cdot)\right] := \{\underline{x} \in \mathbb{R}^n \colon \zeta(\underline{x}) \geq \delta\}$ is a finite union of simply connected compact subsets of $\mathbb{R}^n$.

($\underline{2}$) $C_1\left[\zeta(\cdot, \cdots, \cdot)\right] \neq \emptyset$.

Functions $\zeta(\cdot, \cdots, \cdot)$ obeying the conditions ($\underline{1}$) and ($\underline{2}$) are called *vector-characterizing* functions.

Remark 2: It is important to note that a vector $(x_1^\star, \cdots, x_n^\star)$ of non-precise numbers $x_i^\star$ is not a fuzzy vector. But it is possible to combine the characterizing functions $\xi_i(\cdot)$ of $x_i^\star$, $i = 1, 2, \cdots, n$ in order to obtain

*Corresponding author. Email: R.Viertl@tuwien.ac.at.

the so-called *fuzzy combined sample* $\underline{x}^\star$ whose vector-characterizing function $\zeta(\cdot, \cdots, \cdot)$ is defined by its values $\zeta(x_1, \cdots, x_n)$ for $\forall$ $(x_1, \cdots, x_n) = \underline{x} \in I\!R^n$ in the following way:

$$\zeta(x_1, \cdots, x_n) := \min\left\{\xi_1(x_1), \cdots, \xi_n(x_n)\right\} \qquad \forall \quad (x_1, \cdots, x_n) \in I\!R^n,$$

$\zeta(\cdot, \cdots, \cdot)$ is a vector-characterizing function, because $C_\delta \zeta(\cdot, \cdots, \cdot) = \prod_{i=1}^n C_\delta\left[\xi_i(\cdot)\right]$ for $\forall \quad \delta \in (0; 1]$.

The fuzzy combined sample $\underline{x}^\star$ is the basis for the generalization of Bayesian inference to the situation of fuzzy data.

2 Bayes' Theorem and Fuzzy Data

For continuous parametric stochastic model $X \sim f(\cdot\,|\,\theta)$, $\theta \in \Theta$, a-priori density $\pi(\cdot)$ and data $x_1, \cdots, x_n$ the a-posteriori density $\pi(\cdot \mid x_1, \cdots, x_n)$ of the parameter is obtained by the classical Bayes' theorem

$$\pi(\theta \mid x_1, \cdots, x_n) = \frac{\pi(\theta) \cdot \prod\limits_{i=1}^n f(x_i \mid \theta)}{\int\limits_\Theta \pi(\theta) \cdot \prod\limits_{i=1}^n f(x_i \mid \theta)\, d\theta}, \qquad \forall \quad \theta \in \Theta.$$

Using the notation of the likelihood function

$$\ell(\theta; x_1, \cdots, x_n) = \prod_{i=1}^n f(x_i \mid \theta),$$

Bayes' theorem reads

$$\pi(\theta \mid x_1, \cdots, x_n) \ \propto\ \pi(\theta) \cdot \ell(\theta; x_1, \cdots, x_n), \qquad \forall \quad \theta \in \Theta,$$

where $\propto$ means "proportional to up to a multiplicative constant", i.e.

$$\pi(\theta \mid x_1, \cdots, x_n) = C \cdot \pi(\theta) \cdot \ell(\theta; x_1, \cdots, x_n), \qquad \forall \quad \theta \in \Theta.$$

For fuzzy data $x_1^\star, \cdots, x_n^\star$ the likelihood function has to be generalized. The basis for that is the combined fuzzy sample $\underline{x}^\star$ with its vector-characterizing function $\zeta(\cdots)$. The generalized likelihood function $\ell^\star(\theta; \underline{x}^\star)$ is a fuzzy valued function assigning to every $\theta \in \Theta$ a fuzzy number $\ell^\star(\theta; \underline{x}^\star)$ whose characterizing function is given by application of the so-called *extension principle* from fuzzy set theory: The characterizing function $\psi_{\ell^\star(\theta;\underline{x}^\star)}(\cdot)$ of $\ell^\star(\theta; \underline{x}^\star)$ is obtained by

$$\psi_{\ell^\star(\theta;\underline{x}^\star)}(y) = \left\{ \begin{array}{ll} \sup\left\{\zeta(\underline{x}) \,:\, \ell(\theta;\underline{x}) = y\right\} & \text{if } \exists \ \underline{x} : \ \ell(\theta;\underline{x}) = y \\ 0 & \text{if } \nexists \ \underline{x} : \ \ell(\theta;\underline{x}) = y \end{array} \right\}, \qquad \forall \quad y \in I\!R.$$

Looking at the δ-cuts $C_\delta\left[\psi_{\ell^\star(\theta;\underline{x}^\star)}(\cdot)\right] = \left[\underline{\ell}_\delta(\theta; \underline{x}^\star); \overline{\ell}_\delta(\theta; \underline{x}^\star)\right]$ for varying θ, two classical real valued functions $\underline{\ell}_\delta(\cdot; \underline{x}^\star)$ and $\overline{\ell}_\delta(\cdot; \underline{x}^\star)$ are obtained, called *δ-level functions.*

The generalized (fuzzy) a-posteriori density $\pi^\star(\cdot \mid x_1^\star, \cdots, x_n^\star) = \pi^\star(\cdot \mid \underline{x}^\star)$ is obtained by its δ-level functions $\underline{\pi}_\delta(\cdot \mid \underline{x}^\star)$ and $\overline{\pi}_\delta(\cdot \mid \underline{x}^\star)$ $\forall \quad \delta \in (0; 1]$. These δ-level functions are defined by

$$\left. \begin{array}{l} \overline{\pi}_\delta(\theta \mid \underline{x}^\star) = \dfrac{\pi(\theta) \cdot \overline{\ell}_\delta(\theta; \underline{x}^\star)}{\displaystyle\int_\Theta \pi(\theta) \cdot \frac{\underline{\ell}_\delta(\theta; \underline{x}^\star) + \overline{\ell}_\delta(\theta; \underline{x}^\star)}{2}\, d\theta} \\ \text{and} \\ \underline{\pi}_\delta(\theta \mid \underline{x}^\star) = \dfrac{\pi(\theta) \cdot \underline{\ell}_\delta(\theta; \underline{x}^\star)}{\displaystyle\int_\Theta \pi(\theta) \cdot \frac{\underline{\ell}_\delta(\theta; \underline{x}^\star) + \overline{\ell}_\delta(\theta; \underline{x}^\star)}{2}\, d\theta} \end{array} \right\}, \qquad \forall \quad \left\{ \begin{array}{l} \theta \in \Theta \\ \delta \in (0; 1]. \end{array} \right.$$

Remark 3: $\pi^\star(\cdot \mid \underline{x}^\star)$ is a fuzzy valued function such that all δ-level curves are integrable.

By the sequential updating procedure of Bayesian inference it is necessary to consider fuzzy a-priori densties $\pi^\star(\cdot)$ also.

3 Generalized Bayes' Theorem for Fuzzy A-priori Density and Fuzzy Data

In case of fuzzy a-priori density $\pi^\star(\cdot)$ with δ-level functions $\underline{\pi}_\delta(\cdot)$ and $\overline{\pi}_\delta(\cdot)$ as well as fuzzy data with fuzzy combined sample $\underline{x}^\star$, whose vector-characterizing function is $\zeta(\cdot,\cdots,\cdot)$, Bayes' theorem can be generalized, defining δ-level functions, in the following way:

$$\left.\begin{aligned}
\overline{\pi}_\delta(\theta \mid \underline{x}^\star) &= \frac{\overline{\pi}_\delta(\theta)\cdot\overline{\ell}_\delta(\theta;\underline{x}^\star)}{\int\limits_\Theta \frac{1}{2}\left[\underline{\pi}_\delta(\theta)\cdot\underline{\ell}_\delta(\theta;\underline{x}^\star)+\overline{\pi}_\delta(\theta)\cdot\overline{\ell}_\delta(\theta;\underline{x}^\star)\right]d\theta}\\
&\text{and}\\
\underline{\pi}_\delta(\theta \mid \underline{x}^\star) &= \frac{\underline{\pi}_\delta(\theta)\cdot\underline{\ell}_\delta(\theta;\underline{x}^\star)}{\int\limits_\Theta \frac{1}{2}\left[\underline{\pi}_\delta(\theta)\cdot\underline{\ell}_\delta(\theta;\underline{x}^\star)+\overline{\pi}_\delta(\theta)\cdot\overline{\ell}_\delta(\theta;\underline{x}^\star)\right]d\theta}
\end{aligned}\right\} \quad \forall \quad \left\{\begin{array}{l}\theta\in\Theta\\ \delta\in(0;1].\end{array}\right.$$

Remark 4: By this definition of the δ-level functions $\underline{\pi}_\delta(\cdot \mid \underline{x}^\star)$ and $\overline{\pi}_\delta(\cdot \mid \overline{x}^\star)$ of the fuzzy a-posteriori density the sequentially calculated a-posteriori density $\pi^\star(\cdot \mid \underline{x}_1^\star, \underline{x}_2^\star)$ is the same as the one-step calculated a-posteriori density $\pi^\star(\cdot \mid \underline{x}^\star)$, where $\underline{x}^\star = (\underline{x}_1^\star, \underline{x}_2^\star)$.

Moreover for standard a-priori densities and precise data this concept reduces to the classical Bayes' theorem.

Example 1: For a stochastic quantity X with exponential distribution, i.e. density $f(x \mid \theta) = \frac{1}{\theta}\, e^{-x/\theta}\, I_{(0;\infty)}(x)$ with $\theta \in \Theta = (0;\infty)$, and fuzzy a-priori density $\pi^\star(\cdot)$ on Θ depicted in Figure 1, where $\underline{\pi}_{0^+}(\theta)$ and $\overline{\pi}_{0^+}(\theta)$ are the boundary points of the support of $\pi^\star(\theta)$, some δ-level curves are given.

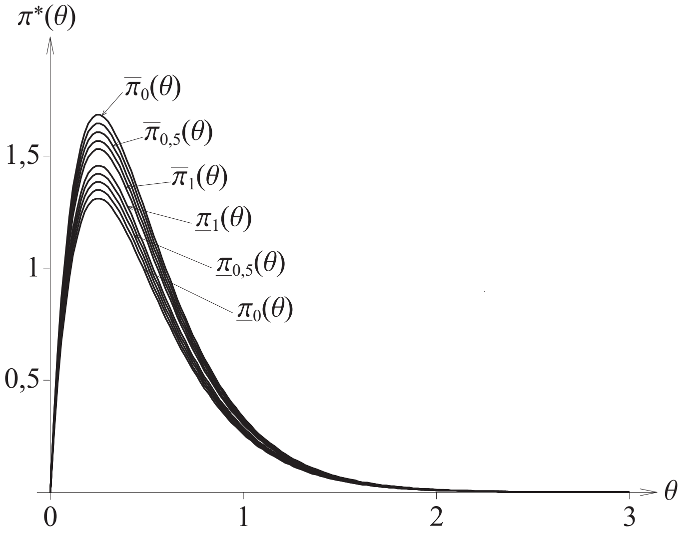

Figure 1: Some δ-level curves of $\pi^\star(\cdot)$

In Figure 2 the characterizing functions of a sample of 8 fuzzy observed life times are depicted. Application of the generalized Bayes' theorem gives the δ-level curves of the fuzzy a-posteriori density $\pi^\star(\cdot \mid x_1^\star, \cdots, x_8^\star)$. The result is depicted in Figure 3.

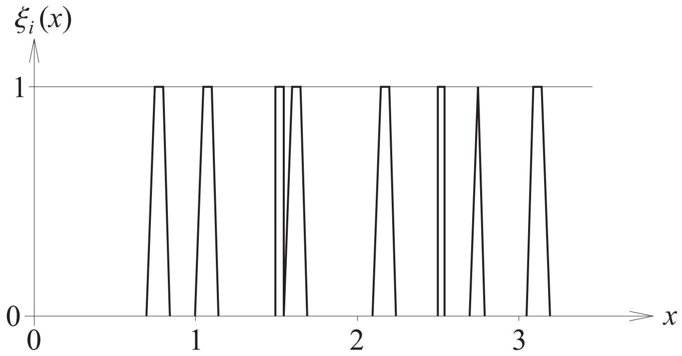

Figure 2: Characterizing functions of a sample of 8 fuzzy life times

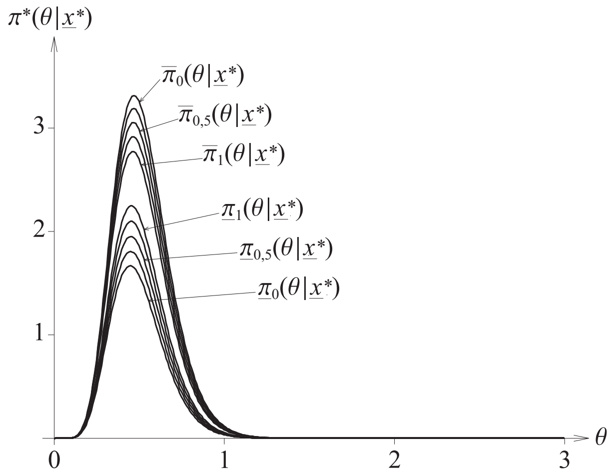

Figure 3: Fuzzy a-posteriori density

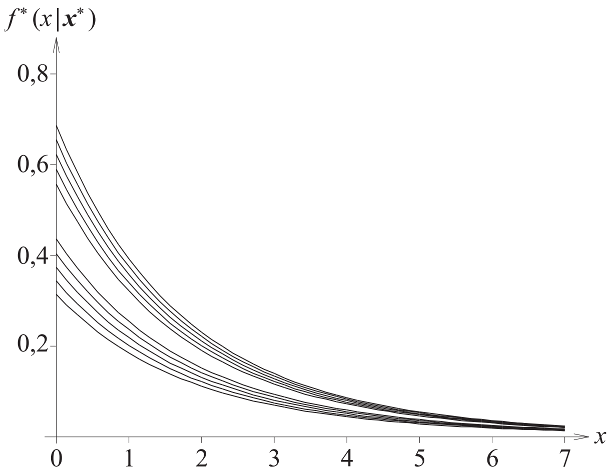

Figure 4: Fuzzy predictive density

4 Predictive Distributions

Let $X \sim f(\cdot \mid \theta)$, $\theta \in \Theta$ be a continuous stochastic model with parametric family of densities $f(\cdot \mid \theta)$, $\theta \in \Theta$. Then the a-posteriori predictive distribution of X conditional on the observed data $x_1, \cdots, x_n$ of X is of interest.

In standard Bayesian inference the so-called *predicitive density* $f(\cdot \mid x_1, \cdots, x_n)$ is obtained as the marginal density of X in $(X, \tilde{\theta})$ based on the a-posteriori density $\pi(\cdot \mid x_1, \cdots, x_n)$, i.e.

$$f(x \mid x_1, \cdots, x_n) = \int_\Theta f(x \mid \theta) \cdot \pi(\theta \mid x_1, \cdots, x_n)\, d\theta \qquad \forall \quad x \in M_X.$$

In case of fuzzy a-posteriori density $\pi^\star(\cdot \mid x_1^\star, \cdots, x_n^\star)$ the resulting predictive density is a fuzzy probability density $f^\star(\cdot \mid x_1^\star, \cdots, x_n^\star)$, whose values are obtained by the generalized integral of fuzzy valued functions:

$$f^\star(x \mid x_1^\star, \cdots, x_n^\star) := ⨍_\Theta f(x \mid \theta) \odot \pi^\star(\theta \mid x_1^\star, \cdots, x_n^\star)\, d\theta \qquad \forall \quad x \in M_X,$$

where $\odot$ denotes the multiplication of a fuzzy valued function by a real number and $⨍$ denotes the integration operator of fuzzy valued functions which is defined via δ-level curves. For details compare [5].

The result of this generalized integration is a fuzzy number. Therefore the generalized predictive density is a fuzzy valued probability density.

In continuation of the example the corresponding fuzzy predictive density is depicted in Figure 4.

5 Conclusion

Basic for Bayesian statistical inference are a-priori distributions and sample data. Both are frequently not precise as is assumed in standard Bayesian inference. Fuzzy a-priori information can be quantified by so-called fuzzy probability densities, and fuzzy data can be modelled by non-precise numbers.

A suitable generalization of Bayes' theorem is presented in the paper which keeps the sequential nature of the Bayesian updating procedure when samples are splitted. Moreover generalized predictive distributions are explained. Further methods of statistical analysis procedures in case of fuzzy data can be found in the German language monograph [5].

References

[1] D'Agostini, G., *Bayesian Reasoning in Data Analysis - A Critical Introduction*, World Scientific Publ., New Jersey, 2003.

[2] Bandemer, G., *Mathematics of Uncertainty - Ideas, Methods, Application Problems*, Springer, Berlin, 2006.

[3] Klir, G., and B. Yuan, *Fuzzy Sets and Fuzzy Logic - Theory and Applications*, Prentice Hall, Upper Saddle River, 1995.

[4] Viertl, R., On the description and analysis of measurement results of continuous quantities, *Kybernetika*, vol.38, no.3, pp.353-362, 2002.

[5] Viertl, R., and D. Hareter, *Beschreibung und Analyse Unscharfer Information*, Springer, Wien, 2006.