

Maximum-Likelihood Principle For Possibility Distributions Viewed As Families Of Probabilities

Mathieu Serrurier,
IRIT,

University of Toulouse III,
31062 Toulouse Cedex 9, France,
E-mail: serrurie@irit.fr

Henri Prade,
IRIT,

University of Toulouse III,
31062 Toulouse Cedex 9, France,
E-mail: prade@irit.fr

Abstract—An acknowledged interpretation of possibility distributions in quantitative possibility theory is in terms of families of probabilities that are upper and lower bounded by the associated possibility and necessity measures. This paper proposes a likelihood function for possibility distributions that agrees with the above-mentioned view of possibility theory in the continuous and in the discrete cases. Especially, we show that, given a set of data following a probability distribution, the optimal possibility distribution with respect to our likelihood function is the distribution obtained as the result of the probability-possibility transformation that obeys the maximal specificity principle. It is also shown that when the optimal distribution is not available, a direct application of this possibilistic likelihood provides more faithful results than approximating the probability distribution and then applying the probability possibility transformation. We detail the particular case of triangular and trapezoidal possibility distributions and we show that any unimodal unknown probability distribution can be faithfully upper approximated by a triangular distribution obtained by optimizing the possibilistic likelihood.

keywords : possibility theory, probability-possibility transformation, maximum-likelihood principle.

I. INTRODUCTION

Possibility theory, based on max-decomposable set-functions, associated with possibility distributions, may have either a qualitative or a quantitative understanding, depending on the nature of the scales used for possibility degrees. Quantitative possibility theory corresponds to the case where the interval $[0,1]$ is a ratio scale. In qualitative possibility theory, only the ordering of the possibility values makes sense. In this paper, we deal with quantitative possibility theory. Quantitative possibility measures can be viewed as upper bound of probabilities. Then, a possibility distribution represents a family of probability distributions [3]. This view was first suggested by Zadeh [10] based on the idea that what is probable must be possible. Following this intuition a probability-possibility transformation has been proposed [5]. This transformation associates a probability distribution with the maximally specific possibility distribution which is such that the possibility of any event is indeed an upper bound of the corresponding probability.

The maximum-likelihood principle is a basis for building a probability distribution from a set of data. Given a set of parameterized probability distributions and a set of data, a

likelihood function is used for optimizing the choice of the parameters in order to determine the best suited distribution with respect to the data. However, this approach supposes that the shape of the distribution is apriori known or that we have a large number of data at our disposal. Moreover, due to the constraints on the probability distributions, induced by their additive nature, there are no simple distribution that allows for the faithful description of any set of data.

There exist different kinds of methods for eliciting possibility distributions from data. For instance, some approaches directly build the possibility distribution on the basis of a proximity relation defined on the universe of the data [4]. G. Mauris proposes a method for constructing a possibility distribution when only very few pieces of data are available (even only one or two) based on probability inequalities [8]. This latter method is justified in the probabilistic view of possibility theory. These methods, how different they are, have in common to build the distributions directly. In this paper, we investigate another road based on the optimization of an appropriate likelihood function. The proposed likelihood function is in agreement with the view underlying the probability-possibility transformation.

The paper is organized as follows. Section 2 provides the necessary background about possibility distributions and their interpretation in terms of families of probabilities. Section 3 presents likelihood measures for probabilities and especially a non-classical one, not based on logarithm, that can be used for approximating unbounded distributions with bounded ones. Sections 4 and 5 respectively deal with the discrete and continuous definitions of possibilistic likelihood and the specific cases of triangular or trapezoidal distributions. Finally, we provide some examples of the construction of a possibility distribution from data using the proposed possibilistic likelihood, and we discuss the usability of the approach in Section 6.

II. BACKGROUND

A. Possibility distribution

Possibility theory, introduced Zadeh [10], was initially created in order to deal with imprecision and uncertainty due

to incomplete information. This kind of uncertainty may not be handled by probability theory, especially when a priori knowledge about the nature of the probability distribution is lacking. A possibility distribution π is a mapping from Ω to $[0, 1]$ (Ω may be a discrete universe, i.e. $\Omega = \{C_1, \dots, C_q\}$, or a continuous one, i.e. $\Omega = \mathbb{R}$). The value $\pi(x)$ is called possibility degree. For any subset of Ω , the possibility measure is defined as follows:

$$\forall A \subseteq \Omega, \Pi(A) = \max\{\pi(x), x \in A\}.$$

If it exists a single value $x \in \Omega$ for which we have $\pi(x) = 1$, the distribution is normalized. We can distinguish two extreme cases of information situations:

- complete knowledge: $\exists x \in \Omega$ such as $\pi(x) = 1$ and $\forall y \in \Omega, y \neq x, \pi(y) = 0$
- total ignorance: $\forall x \in \Omega, \pi(x) = 1$.

The necessity is the dual measure of the possibility measure. We have:

$$\forall A \subseteq \Omega, N(A) = 1 - \Pi(\bar{A}).$$

Finally, α -cuts of π are subsets of Ω such that:

$$A_\alpha = \{x \in \Omega, \pi(x) \geq \alpha\}.$$

It can be checked that, if $\Omega = \mathbb{R}$ and the distribution is continuous and normalized, we have $\forall \alpha \in [0, 1], \Pi(A_\alpha) = 1$ and $N(A_\alpha) = 1 - \alpha$.

B. Possibility distribution as a family of probability distributions

Possibility distributions have two types of interpretations. The first one, that is related to fuzzy set theory, is the description of gradual properties. For instance, the definition of linguistic expressions such that “long”, “old” or “expensive” does not refer to a specific value, but to a set of possible values in a particular context. For instance, a possibility distribution may describe the concept “expensive” for an house in a given area. In such a case, each price will be associated with a possibility degree which quantifies how much this price is typical with respect to this particular use of the concept “expensive”.

Another type of interpretation is to consider a possibility distribution as a family of probability distributions (see [1] for an overview). Thus, a possibility distribution π will represent the family of the probability distributions for which the measure of each subset of Ω will be bounded by its necessity and its possibility measures. More formally, if $\mathcal{P}$ is the set of all probability distributions defined on Ω , the family of probability distributions $\mathcal{P}_\pi$ associated with π is defined as follows:

$$\mathcal{P}_\pi = \{p \in \mathcal{P}, \forall A \in \Omega, N(A) \leq P(A) \leq \Pi(A)\}. \quad (1)$$

In this scope, the situation of total ignorance corresponds to the case where all probability distributions are possible. This type of ignorance can not be described by a single probability distribution. The case of complete knowledge corresponds to the case where only one value is possible and then where there are no randomness nor imprecision. Obviously, if π is not

normalized, $\mathcal{P}_\pi$ is empty (since then $N(A) \leq \Pi(A)$ is no longer guaranteed for any A).

When $\Omega = \mathbb{R}$, this family of probability distributions can also be described in terms of confidence intervals. Given a probability distribution p , a confidence interval I_α is a subset of Ω such as $P(I_\alpha) = \alpha$. We define I_α^* , also referred as quantile, as the smallest confidence interval with probability measure equal to α . Thus, when π is continuous, an alternative to the equation (1) is:

$$\mathcal{P}_\pi = \{p \in \mathcal{P}, \forall I_\alpha^* \in \Omega, I_\alpha^* \subseteq A_{1-\alpha}\} \quad (2)$$

where $A_{1-\alpha}$ is the $(1 - \alpha)$ -cut of π . Thus, the possibility distribution π contains the probability distributions for which the confidence intervals associated to the α 's are bounded by the $(1 - \alpha)$ -cuts.

C. Probability to possibility transformation

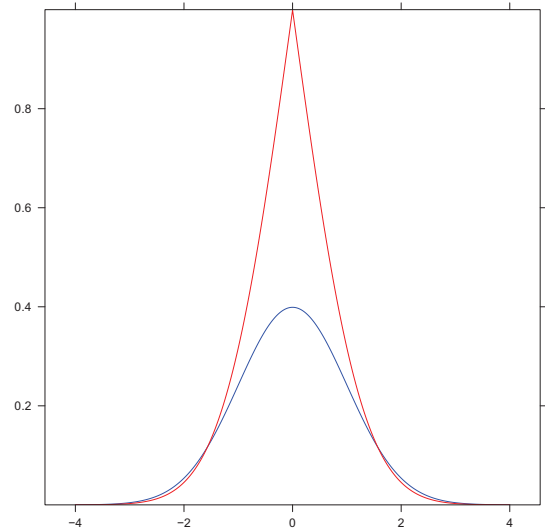


Fig. 1. probability to possibility transformation of a Gaussian distribution

According to this probabilistic interpretation, a method from transforming probability distributions into possibility distributions has been proposed in [5]. The idea behind this proposal is to consider the most informative possibility distribution, i.e. the tightest one (i.e. the smallest one in each point) that contains the probability distribution. Let us consider a probability distribution p , the possibility distribution π^* is defined in the following way:

$$\forall x \in \Omega, \pi^*(x) = \max_{\alpha, x \in I_\alpha} (1 - \alpha). \quad (3)$$

Then, in the spirit of equation (2), given p and its transformation π^* we have:

$$A_{1-\alpha}^* = I_\alpha^*$$

where $A_{1-\alpha}^*$ is the $(1 - \alpha)$ -cut of π^* . Thus, if p has a finite number of modes, π^* is the possibility distribution for which each $(1 - \alpha)$ -cut corresponds to the quantile of p . When p is

unimodal, the unique value x such that $\pi^*(x) = 1$ is the mode of p .

III. PROBABILISTIC LIKELIHOOD MEASURE

Likelihood measures have been introduced in order to evaluate the adequateness of a probability distribution with respect to a set of data. Let us consider a set of data $X = \{x_1, \dots, x_n\}$ belonging to a discrete universe $\Omega = \{C_1, \dots, C_q\}$. We name $\alpha_1, \dots, \alpha_q$ the frequency of the elements of X that belong respectively to $\{C_1, \dots, C_q\}$. Given a probability distribution p on the discrete space $\Omega = \{C_1, \dots, C_q\}$, we define $p_1, \dots, p_q$ the probability of belonging to the unitary element of Ω , i.e. $p(x \in C_i) = p_i$. The values $p_1, \dots, p_q$ entirely define p , and are then the parameters of p . The maximum-likelihood principle is used for estimating the parameters of a probability distribution. In the continuous case, the shape of the distribution has to be known, and the parameters are obtained through an optimization procedure with respect to the likelihood measure. In the discrete case, the parameters are $p_1, \dots, p_q$, and obey the constraint $\sum_{i=1}^q p_i = 1$. The most used likelihood function is the logarithmic-based likelihood defined as follows (under the strict constraint $\sum_{i=1}^q p_i = 1$):

$$\mathcal{L}_{\log}(p|x_1, \dots, x_n) = - \sum_{i=1}^n \log(p(x_i))$$

or, when considering frequency directly

$$\mathcal{L}_{\log}(p|x_1, \dots, x_n) = - \sum_{i=1}^q \alpha_i \log(p_i).$$

It is equivalent to compute the joint probability of the elements of x with respect to p . This definition of the likelihood has a strong limitation, since it gives a very high weight to the error when probability is very low. This is especially true when Ω is continuous. Since $\mathcal{L}_{\log}$ is not defined when $p(x_i) = 0$, an unbounded density cannot be approximated by a bounded one by optimization of $\mathcal{L}_{\log}$. We propose another likelihood function that overcomes these limitations:

$$\mathcal{L}_{surf}(p|x_1, \dots, x_n) = \left(\sum_{i=1}^n p(x_i) \right) - \frac{1}{2} * \sum_{i=1}^q p_i^2$$

or, when considering frequency directly

$$\mathcal{L}_{surf}(p|x_1, \dots, x_n) = \left(\sum_{i=1}^q \alpha_i * p_i \right) - \frac{1}{2} * \sum_{i=1}^p p_i^2.$$

Roughly speaking, $\mathcal{L}_{surf}$ favors the probability distributions that are close to the optimal one in terms of surface. Thus, when, Ω is continuous, it allows for the approximation of unbounded densities by bounded ones.

Proposition 1: Given a set of data $X = \{x_1, \dots, x_n\}$ belonging to a discrete universe $\Omega = \{C_1, \dots, C_q\}$, we have

$$\begin{aligned} & \operatorname{argmax}_{p \in \mathcal{P}} (\mathcal{L}_{\log}(p|x_1, \dots, x_n)) \\ &= \operatorname{argmax}_{p \in \mathcal{P}} (\mathcal{L}_{surf}(p|x_1, \dots, x_n)). \end{aligned}$$

Proof: Let $p_{\log} = \operatorname{argmax}_{p \in \mathcal{P}} (\mathcal{L}_{\log}(p|x_1, \dots, x_n))$ be the optimal probability distribution given X . This distribution is such as the probability of an event C_i is equal to the frequency of element of X in C_i , i.e. $p_{\log}(x \in C_i) = p_i = \alpha_i$. We now look for the probability distribution p_{surf} that maximizes $\mathcal{L}_{surf}$. We have:

$$\forall i = 1 \dots q, \frac{\delta \mathcal{L}_{surf}(p|x_1, \dots, x_n)}{\delta p_i} = \alpha_i - p_i$$

thus

$$\forall i = 1 \dots q, \frac{\delta \mathcal{L}_{surf}(p|x_1, \dots, x_n)}{\delta p_i} = 0 \Leftrightarrow p_i = \alpha_i.$$

Since the derivative of $\mathcal{L}_{surf}(p|x_1, \dots, x_n)$ with respect to p_i (the parameters of p) does not depend on the other parameters $p_j, j \neq i$, we obtain $p_{surf}(x \in C_i) = p_i = \alpha_i$. Thus $p_{surf} = p_{\log}$ ■

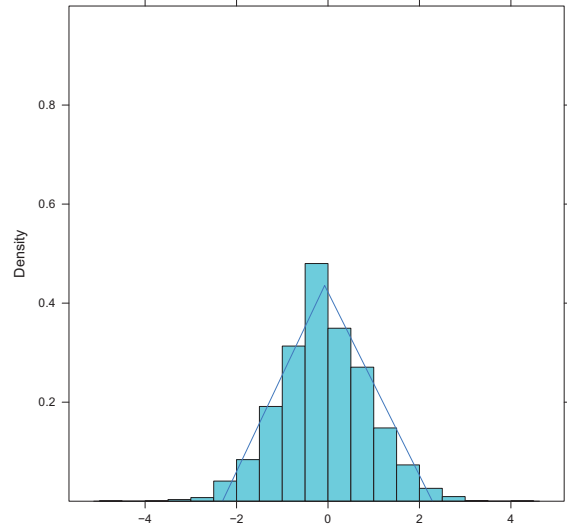


Fig. 2. Approximation of a Gaussian density with a triangular density by maximization of $\mathcal{L}_{surf}$

This proposition shows that, given X , the probability distribution that maximizes $\mathcal{L}_{\log}$ is the same as the one that maximizes $\mathcal{L}_{surf}$.

$\mathcal{L}_{surf}$ can be extended to the case of $\Omega = \mathbb{R}$ in the following way:

$$\mathcal{L}_{surf}(p|x_1, \dots, x_n) = \left(\sum_{i=1}^n p(x_i) \right) - \frac{1}{2} * \int_{\mathbb{R}} p(t)^2 dt.$$

Proposition 1 remains true when $\Omega = \mathbb{R}$. Figure 2 shows the triangular density that maximizes $\mathcal{L}_{surf}$ given a set of 3000 values that are generated by a Gaussian distribution with mean equal to 0 and standard derivation equal to 1. This triangular probability distribution is parametrized by m ,

l and r which are respectively the mode, the left and the right spread of the probability distribution. We have used an optimization procedure in order to find the parameters that maximizes $\mathcal{L}_{surf}$. What we obtain is the triangular probability distribution that shares the maximum of surface with the histogram computed from the sample set. If we have used $\mathcal{L}_{log}$, the support of the triangle would have grow to infinity as the number of values increases.

IV. POSSIBILISTIC LIKELIHOOD: DISCRETE CASE

In this section we show how to use the $\mathcal{L}_{surf}$ likelihood function in order to define a likelihood function for a possibility distribution that supports the interpretation of a possibility distribution in terms of a family of probability distributions. Roughly speaking, the idea is to have a likelihood function that is maximal for the result of the probability-possibility transformation of the distribution and that favors the approximations of the distribution that share the maximum of surface with the optimal one (as for $\mathcal{L}_{surf}$ in the probabilistic case). We first consider the case of a discrete universe, i.e. $\Omega = \{C_1, \dots, C_q\}$.

We now assume that the frequencies of examples in the different classes C_i are put in increasing order ($\alpha_1 \geq \dots \geq \alpha_q$). In this case, the probability distribution p^* that maximizes the likelihood is such that $p(x \in C_i) = p_i = \alpha_i$. In the following, given a possibility distribution π , we note π_i the value $\pi(x \in C_i)$. It has been shown in [5] that the transformation of p^* into a possibility distribution π^* that satisfies the maximal specificity requirement, is:

$$\forall i \in \{1, \dots, q\}, \pi_i^* = \sum_{j=i}^q \alpha_j. \quad (4)$$

This possibility distribution is one of the cumulated functions of p^* . It is worth noticing that it is the tightest one. What we expect from possibility likelihood is that the maximum of this function is reached for π^* . In the following, we assume that $\pi_1 \geq \dots \geq \pi_q$ (and not necessarily $\alpha_1 \geq \dots \geq \alpha_q$). We propose the following function:

$$\begin{aligned} \mathcal{L}_{pos}(\pi|x_1, \dots, x_n) = & \left(\sum_{i=1}^q (1 - \pi_i) * \left(1 - \left(\sum_{j=i}^q \alpha_j \right) \right) \right) \\ & - \sum_{i=1}^q \frac{(1 - \pi_i)^2}{2} \end{aligned} \quad (5)$$

which can be rewritten as this:

$$\begin{aligned} \mathcal{L}_{pos}(\pi|x_1, \dots, x_n) = & \sum_{i=1}^q \left(-\alpha_i * \sum_{j=1}^i (1 - \pi_j) \right) \\ & - \sum_{i=1}^q \frac{(1 - \pi_i)^2}{2} \\ & + \sum_{i=1}^q (1 - \pi_i) \end{aligned} \quad (6)$$

If we only consider one piece of data, we obtain:

$$\begin{aligned} \mathcal{L}_{pos}(\pi|x) = & \sum_{j=1}^{i|x \in C_i} (1 - \pi_j) \\ & - C * \sum_{i=1}^q \frac{(1 - \pi_i)^2}{2} \\ & + C * \sum_{i=1}^q (1 - \pi_i) \end{aligned} \quad (7)$$

where C is a constant (usually $\frac{1}{n}$, where n is the number of independent data considered). The rationale behind this is to evaluate the cumulated distribution in the spirit of $\mathcal{L}_{surf}$. Thus, the terms $((1 - \pi_i) * (1 - (\sum_{j=i}^q \alpha_j))) - \frac{(1 - \pi_i)^2}{2}$ correspond to the evaluation of $\mathcal{L}_{surf}$ for the set $\overline{C_i} \cup \dots \cup \overline{C_q}$. Note that, if you consider $C_i \cup \dots \cup C_q$ instead of $\overline{C_i} \cup \dots \cup \overline{C_q}$, what is obtained is the largest cumulative distribution instead of the tightest one.

Proposition 2: Given a set of data $X = \{x_1, \dots, x_n\}$ belonging to a discrete universe $\Omega = \{C_1, \dots, C_q\}$, the possibility distribution π^* that maximizes the function $\mathcal{L}_{pos}$ is the transformation of the probability distribution p^* , according to equation (4), which maximizes $\mathcal{L}_{surf}$ ($\forall i \in \{1, \dots, q\}, p_i^* = \alpha_i$).

Proof: We look for the probability distribution π^* that maximizes $\mathcal{L}_{surf}$.

We have:

$$\forall i = 1 \dots q, \frac{\delta \mathcal{L}_{pos}(\pi|x_1, \dots, x_n)}{\delta \pi_i} = -1 + \sum_{j=i}^q \alpha_j - \pi_i + 1$$

thus

$$\forall i = 1 \dots q, \frac{\delta \mathcal{L}_{pos}(\pi|x_1, \dots, x_n)}{\delta \pi_i} = 0 \Leftrightarrow p_i = \sum_{j=i}^q \alpha_j.$$

Since the derivative of $\mathcal{L}_{pos}(\pi|x_1, \dots, x_n)$ with respect to π_i (the parameters of π) does not depend on the other operator π_j , $j \neq i$, we obtain $\pi_i^* = p_i = \sum_{j=i}^q \alpha_j$ which exactly corresponds to Equation (4). ■

This proposition shows that $\mathcal{L}_{pos}$ is an acceptable likelihood function for possibility distributions viewed as families of probabilities. As for $\mathcal{L}_{surf}$ the likelihood depends on the surface shared between the considered possibility distribution and the optimal one.

It is worth noticing that, when optimal distributions can only be approximated, finding the best approximation with respect to $\mathcal{L}_{pos}$ is not equivalent to find the best probability approximation with respect to $\mathcal{L}_{surf}$ and then turn it into a possibility distribution. For instance, we consider X that leads to the frequency $\alpha_1 = 0.5, \alpha_2 = 0.3, \alpha_3 = 0.2$. We know require that $p_3 = 0$ and $\pi_3 = 0$. In this context, the optimal p with respect to $\mathcal{L}_{surf}$ ($\mathcal{L}_{log}$ is not applicable here) is $p_1 = 0.6, p_2 = 0.4, p_3 = 0$. The optimal π with respect to $\mathcal{L}_{pos}$ is

$\pi_1 = 1, \pi_2 = 0.5, \pi_3 = 0$. The transformation π' of p is $\pi'_1 = 1, \pi'_2 = 0.4, \pi'_3 = 0$. We observe that π' is different than π and that π is a better approximation of the optimal possibility distribution.

This result is fundamental since it illustrates that using a probabilistic likelihood and then the probability-possibility transformation is not an effective approach for constructing a possibility distribution from data. The maximization of $\mathcal{L}_{pos}$ is more adapted in this scope.

V. POSSIBILISTIC LIKELIHOOD: CONTINUOUS CASE

We now consider the continuous case where $\Omega = \mathbb{R}$.

A. Definition

In the continuous case, the assumption of the values put in an increasing order of π is naturally replaced by the consideration of α -cuts. We adapt the equation (6) as follows:

$$\begin{aligned} \mathcal{L}_{pos}(\pi|x_1, \dots, x_n) = & - \left(\sum_{i=1}^n \int_{A_{\pi(x_i)}} (1 - \pi(t)) dt \right) \\ & - \int_{\mathbb{R}} \frac{(1 - \pi(t))^2}{2} dt \\ & + \int_{\mathbb{R}} (1 - \pi(t)) dt \end{aligned} \quad (8)$$

where $A_{\pi(x_i)}$ is the $\pi(x_i)$ -cut of π . If we only consider one piece of data, we obtain:

$$\begin{aligned} \mathcal{L}_{pos}(\pi|x) = & - \int_{A_{\pi(x)}} (1 - \pi(t)) dt \\ & + C * \left(- \int_{\mathbb{R}} \frac{(1 - \pi(t))^2}{2} dt + \int_{\mathbb{R}} (1 - \pi(t)) dt \right) \end{aligned} \quad (9)$$

where C is a constant (usually $\frac{1}{n}$, where n is the number of independent data considered). Property 2 remains true in the continuous case. The possibilistic counterpart of likelihood being defined, we will now consider the particular cases of triangular and trapezoidal distributions.

B. Triangular distribution

We define a triangular possibility distribution as the triple $\pi_{tri} = (m, l, r)$ where m is the mode of the triangle and l and r the left and the right spread respectively. Since the 0-cut is infinite for a triangular distribution, we assume that X is bounded and has a maximal size equal to the constant MAX_{size} .

We consider a piece of data $x \in X$. We note $\mu = \pi_{tri}(x)$ the possibility degree of x and $[a, b]$ the μ -cut of π_{tri} . There are two cases for the term that depends on $\pi_{tri}(x)$ in (9). We

consider the case of $x \in]m - l, m + r[$. We have:

$$\begin{aligned} MemSurf(\pi_{tri}|x) = & - \int_{A_{\pi_{tri}(x)}} (1 - \pi_{tri}(t)) dt \\ = & - \int_a^b (1 - \pi_{tri}(t)) dt \\ = & - \int_a^m (1 - \pi_{tri}(t)) dt + \int_m^b (1 - \pi_{tri}(t)) dt \\ = & - \int_a^m \left(1 - \left(1 - \frac{m-t}{l}\right)\right) dt - \int_m^b \left(1 - \left(1 - \frac{t-m}{r}\right)\right) dt \\ = & - \int_a^m \frac{m-t}{l} dt - \int_m^b \frac{t-m}{r} dt \\ = & -(1-\mu)^2 * \frac{l}{2} - (1-\mu)^2 * \frac{r}{2} \\ = & -(1-\mu)^2 * \frac{l+r}{2}. \end{aligned}$$

In the case of $x \notin]m-l, m+r[$, with the bounding assumption, we obtain:

$$MemSurf(\pi_{tri}|x) = -MAX_{size} + \frac{l+r}{2}.$$

Note that a more flexible approach on the bounding can be used, for instance by considering that the weight MAX_{size} depends on the distance between x and the triangle. The other part of the equation (9) is computed such as:

$$\begin{aligned} & - \int_{\mathbb{R}} \frac{(1 - \pi_{tri}(t))^2}{2} dt + \int_{\mathbb{R}} (1 - \pi(t)) dt = \\ & \int_{m-l}^{m+r} - \frac{(1 - \pi_{tri}(t))^2}{2} + (1 - \pi_{tri}(t)) dt \\ & + \int_{R/[m_l, m+r]} - \frac{(1 - \pi_{tri}(t))^2}{2} + (1 - \pi_{tri}(t)) dt = \\ & - \frac{l+r}{6} - \frac{MAX_{size}}{2}. \end{aligned}$$

The terms MAX_{size} depends neither on π_{tri} and nor on x and can then be omitted. Finally, we obtain:

$$\mathcal{L}_{pos}(\pi_{tri}|x) = MemSurf(\pi_{tri}|x) - C * \frac{l+r}{6}. \quad (10)$$

C. trapezoidal distribution

We define a trapezoidal possibility distribution as the quadruple $\pi_{trap} = (a, b, c, d)$ where $[a, d]$ and $[b, c]$ are respectively the support and the core of the distribution. As in the triangular case, we assume that X is bounded and has a maximal size equal to the constant MAX_{size} . We note $\mu = \pi_{trap}(x)$ the possibility degree of x . There are three cases for the term that depends on $\pi_{tri}(x)$ in Equation (9).

- $x \in]b, c[$:

$$MemSurf(\pi_{trap}|x) = 0$$

- $x \in]a, b[$ or $x \in]c, d[$:

$$MemSurf(\pi_{trap}|x) = -(1-\mu)^2 * \frac{b-a+d-c}{2}$$

- $x \notin]a, d[$:

$$MemSurf(\pi_{trap}|x) = -MAX_{size} + \frac{-b - a + d + c}{2}.$$

The other part of the equation (9) is computed such as:

$$-\int_{\mathbb{R}} \frac{(1 - \pi_{trap}(t))^2}{2} dt + \int_{\mathbb{R}} (1 - \pi(t)) dt = -\frac{b - a + d - c}{6} - \frac{c - b}{2} - \frac{MAX_{size}}{2}.$$

As in the triangular case, the terms $-\frac{MAX_{size}}{2}$ can be omitted. Finally, we obtain:

$$\mathcal{L}_{pos}(\pi_{trap}|x) = MemSurf(\pi_{tri}|x) - C * \left(\frac{b - a + d - c}{6} - \frac{c - b}{2} \right). \quad (11)$$

Note that, having the likelihood function defined for triangular (resp. trapezoidal) distribution, we can obtain the optimal triangular (resp. trapezoidal) by finding the parameters that maximizes $\mathcal{L}_{pos}$. Since this problem cannot be solved analytically, we have to use a meta heuristic such as simulated annealing [7] or particle swarm optimization [6].

VI. ILLUSTRATIONS AND DISCUSSION

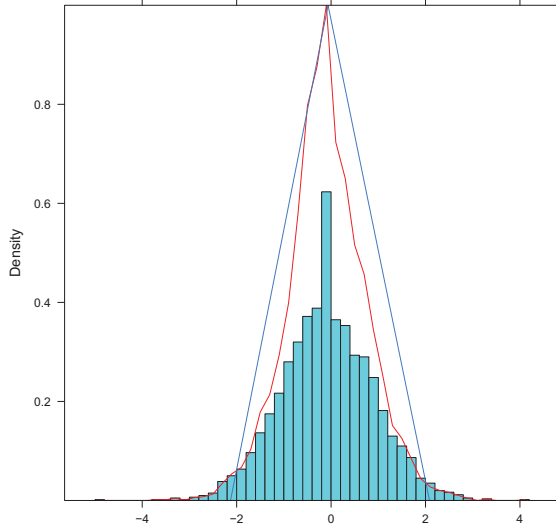


Fig. 3. Possibility transformation that maximizes $\mathcal{L}_{pos}$ with respect to a set of data that follows a Gaussian distribution

In this section, we will use possibilistic likelihood in order to build a possibility distribution from a set of data. Figure 3 represents the triangular possibility distribution obtained from a set of 3000 data that obey a Gaussian distribution with a mean equal 0 and a standard deviation equals to 1. The possibility distribution obtained from a probability-possibility transformation of the histogram is in red. It corresponds to the optimal possibility distribution with respect to data. The triangular possibility distribution that maximizes $\mathcal{L}_{pos}$ is in blue. We can first notice that the mode

of the triangular distribution corresponds to the mode of the Gaussian distribution. The triangular distribution is very close to the optimal one and dominates it on the majority of the data. The support contains 96% of the data. Figure 4

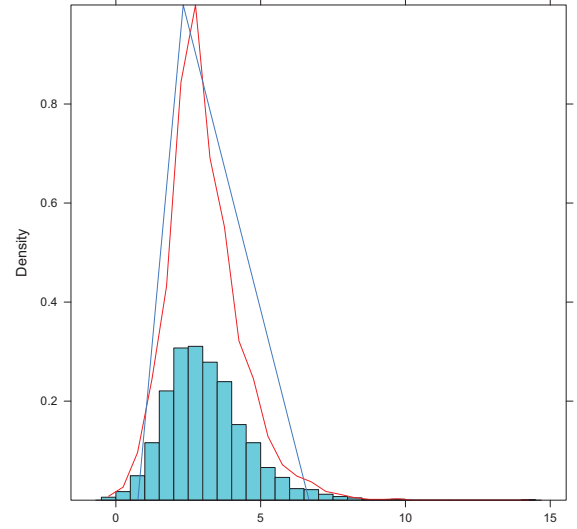


Fig. 4. Possibility transformation that maximizes $\mathcal{L}_{pos}$ with respect to a set of data that follow a Student distribution

represents the triangular possibility distribution obtained from a set of 3000 data that follow a non central student distribution with 5 degrees of freedom. Although the distribution is not symmetric, we can observe similar results to the one observed with the Gaussian distribution. The support contains 96% of the data. In Figure 5, the distribution is multi modal. Here

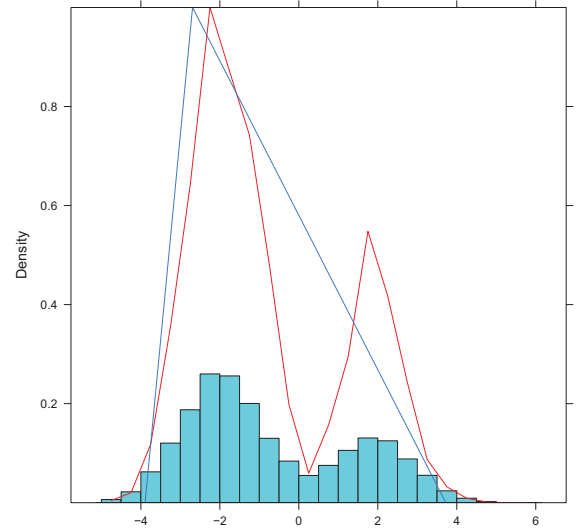


Fig. 5. Possibility transformation that maximizes $\mathcal{L}_{pos}$ with respect to a set of data that follows a multimodal distribution

we can observe that the mode of the triangular distribution is the highest mode of the probability distribution. However, it underestimates the second mode. The support still contains 96% of the data.

These results illustrates the following property, proved in [2]:

Proposition 3: The triangular symmetric possibility distribution of support $[x_1, x_2]$ and of mode $\frac{x_1+x_2}{2}$ is the least upper bound of all the possibility transforms of symmetric probability distributions of support $[x_1, x_2]$ and of mode $\frac{x_1+x_2}{2}$.

This proposition shows that triangular distributions may approximate any possibility transformation of bounded symmetric unimodal distributions. We have shown in Section 4 that approximating a possibility distribution by maximizing $\mathcal{L}_{pos}$ is more efficient than approximating the probability distribution and then turning it into a possibility distribution. It validates the use of possibilistic likelihood for building triangular distributions from data. Even Proposition 3 is not always true if the distribution is asymmetric, triangular possibilistic distribution performs well in general for any type of unimodal distribution. Thus, the use of triangular possibility distributions and of $\mathcal{L}_{pos}$ is a good alternative for bounding densities when the shape of the probability distribution is not known. Results are less good if the distribution is multimodal, but the mode of the triangular possibilistic distribution always identifies the highest mode of the probability distribution. Note that we do not present a result with a trapezoidal distribution because triangular distributions (which are a particular case of trapezoidal distributions) are always obtained.

VII. CONCLUSION

In this paper we have proposed a definition of possibilistic likelihood that agrees with the view of possibility distributions as families of probability distributions and with the probability-possibility transformation based on the maximum-specificity principle. This function can be used for evaluating and comparing a possibility distribution with respect to data or can be embedded into an optimization process in order to build a possibility distribution from a set of data. We have defined the possibilistic likelihood function both for discrete and for continuous universes. In the continuous case, we have shown that, in spite of its sophisticated definition, it is easy to compute for triangular and trapezoid distributions. Lastly, the good properties of triangular possibility distributions for bounding unimodal probability distributions make the building of possibility distributions by optimization of $\mathcal{L}_{pos}$ a good approach when no a priori information on the type of distribution is available. This optimization approach may be interesting in a machine learning context, for instance for Bayesian learning or for a k-nearest neighbor approach where distributions have to be computed from data. It can be also used as quality function for imprecise regression [9].

In the future, it may be interesting to define a possibilistic likelihood that allows us to obtain genuine trapezoidal distributions. Moreover, the proposed approach does not take into

account the quantity of data available. Indeed, if only one value is available, the best possibility distribution is a Dirac, and there will be no difference with the case where a thousand of identical values are available. Less data should lead to less specific distributions. This should be taken into account in representations.

REFERENCES

- [1] D. Dubois. Possibility theory and statistical reasoning. *Computational Statistics and Data Analysis*, 51:47–69, 2006.
- [2] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10:273–297, 2004.
- [3] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.
- [4] D. Dubois and H. Prade. On data summarization with fuzzy sets. In *Proc. of the 5th Inter. Fuzzy Systems Assoc. World Congress (IFSA'93)*, Seoul, pages 465–468, 1993.
- [5] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In *Proceedings of Fourth IFSA Conference*, pages 103–112. Kluwer Academic Publ, 1993.
- [6] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, pages 1942–1948, 1995.
- [7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
- [8] G. Mauris. Inferring a possibility distribution from very few measurements. In *Soft Methods for Handling Variability and Imprecision*, volume 48 of *Advances in Soft Computing*, pages 92–99. Springer Berlin / Heidelberg, 2008.
- [9] M. Serrurier and H. Prade. A general framework for imprecise regression. In *FUZZ-IEEE 07, London*, pages 1597–1602, 2007.
- [10] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–25, 1978.