

On Data-Based Estimation of Possibility Distributions

Dominik Hose and Michael Hanss

Institute of Engineering and Computational Mechanics, University of Stuttgart,
Pfaffenwaldring 9, 70569 Stuttgart, Germany
{dominik.hose, michael.hanss}@itm.uni-stuttgart.de

November 20, 2019

Preprint submitted to Fuzzy Sets and Systems on October 14, 2019.

Abstract

In this paper, we show how a possibilistic description of uncertainty arises very naturally in statistical data analysis. In combination with recent results in inverse uncertainty propagation and the consistent aggregation of marginal possibility distributions, this estimation procedure enables a very general approach to possibilistic identification problems in the framework of imprecise probabilities, i.e. the non-parametric estimation of possibility distributions of uncertain variables from data with a clear interpretation.

Keywords: Possibilistic Statistics, Probability-Possibility Consistency, Fuzzy Membership Estimation, Possibility Distribution Estimation, Fuzzy Measurement Equations

1 Introduction

Frequently, when discussing ideas related to fuzzy set theory with researchers and practitioners working in uncertainty quantification, the question arises how the shape of the membership function ought to be assessed. Besides the widely accepted, yet seldom convincing, answer of so-called 'expert knowledge', other options can usually not be provided. Especially, the case of inferring membership functions from data, which may be called 'possibilistic statistics', has been addressed insufficiently and only selectively by the scientific community in the past.

For instance, the idea of inferring possibilistic distributions from measurements appears in [1], where the inputs and outputs are assembled into fuzzy data and afterwards propagated backwards by means of (an approximate) inverse fuzzy arithmetic. The same idea is presented in [2], where the output data is fuzzified in a different manner. However, both approaches lack a convincing explanation for the choice of the shape of the membership functions. On the other hand, Serrurier and Prade develop a maximum likelihood principle in fuzzy set theory [3] where the membership function is essentially the rescaled likelihood function of the data which in the authors view does not draw upon the whole expressiveness of possibility distributions. The construction of membership functions from for interval- or fuzzy-valued data, refer e.g. to [4, 5], shall not be discussed here, as we concentrate on the case of crisp data.

We argue that by viewing fuzzy membership functions as possibility densities for a description of imprecise probabilities, a sensible estimation procedure arises very naturally and may be derived in a theoretically sound manner. A notable paper in this line of thought is that of Masson and Denoeux [6] who infer imprecise probability distributions of discrete variables from samples only and gather them in possibility distributions.

The remainder of this paper is organized as follows. Section 2 briefly summarizes the framework of possibility theory. In Section 3, this framework is discussed with respect to its capabilities of representing imprecise probability distributions. In accordance with the view of possibility distributions as consonant random sets [7], Section 4 is intended to review some basic families of nested sets, and Section 5 discusses the estimation of membership functions thereon. This is the core of this contribution where we provide probabilistic reliability guarantees for the consistency of the obtained possibility distributions. Subsequently, Section 6 is concerned with the inclusion of precise and imprecise measurement models and, finally, in Section 7 an efficient algorithm for the computation of the main results is given. An exemplary application of the main results in Section 8 is followed by a concluding discussion in Section 9.

The examples throughout the paper are intended to make the reader familiar with the theory, show the simplicity of the involved computations, and contribute to the understanding of this paper.

2 Possibilistic Representation of Uncertainty

We will briefly recount the basic terminology of possibility theory here. A set-valued function $\Pi : 2^\Omega \rightarrow [0, 1]$ defined on the power set 2^Ω of the universe of discourse Ω may be called a possibility measure if it satisfies three conditions similar to those of a probability measure. Explicitly, $\Pi(\emptyset) = 0$, $\Pi(\Omega) = 1$ and $\Pi(\bigcup_k U_k) = \sup_k \Pi(U_k)$ for countable collections of sets $U_k \subseteq \Omega$ have to be fulfilled. The dual necessity measure is defined by $N(U) = 1 - \Pi(\Omega \setminus U)$ for all $U \subseteq \Omega$. In particular, the three conditions for a possibility measure are equivalent to requiring that the necessity measure fulfills $N(\emptyset) = 0$, $N(\Omega) = 1$ and $N(\bigcap_k U_k) = \inf_k N(U_k)$ for countable collections of sets $U_k \subseteq \Omega$. Both descriptions are equivalent. An (\mathcal{X} -valued) uncertain (random, fuzzy, etc.) variable $X : \Omega \rightarrow \mathcal{X}$ possesses a possibility distribution $\Pi_X(U) = \Pi(X^{-1}(U))$ and a possibility density $\pi_X : \mathcal{X} \rightarrow [0, 1]$ sufficing $\Pi_X(U) = \sup_{x \in U} \pi_X(x)$ for all $U \subseteq \mathcal{X}$. This implies $\pi_X(x) = \Pi_X(\{x\})$ for $x \in \mathcal{X}$. The sub-/superlevel sets are defined as $\mathcal{L}_{\Pi_X}^{\alpha^-} = \{x \in \mathcal{X} : \pi_X(x) \leq \alpha\}$ and $\mathcal{L}_{\Pi_X}^{\alpha^+} = \{x \in \mathcal{X} : \pi_X(x) \geq \alpha\}$, respectively. The superlevel sets are also known as the α -cuts. Below, $\mathcal{S}(\mathcal{X})$ denotes a σ -field on \mathcal{X} . In practical applications, typically $\mathcal{X} \subseteq \mathbb{R}^N$, and $\mathcal{S}(\mathcal{X})$ is the corresponding Borel σ -algebra.

3 Possibilities as Imprecise Probabilities

Possibility theory offers a general framework for the analysis of imprecisely defined probabilities, i.e. for the analysis of sets of probability distributions. The fundamental principle of probability-possibility consistency was first introduced by Dubois and Prade in [8]. It states that a probability measure P_X and a possibility measure Π_X are consistent if

$$P_X(U) \leq \Pi_X(U) \quad \forall U \in \mathcal{S}(\mathcal{X}). \quad (1)$$

This is equivalent to saying that the corresponding necessity measure N_X and P_X are consistent, i.e. $N_X(U) \leq P_X(U)$ for all $U \in \mathcal{S}(\mathcal{X})$, which follows from $P_X(\mathcal{X} \setminus U) \leq \Pi_X(\mathcal{X} \setminus U)$. However, in order to show consistency one does not need to check this condition for all possible events. Instead, one can infer that it suffices to check the condition for the super- or sublevel sets as shown in [9].

Proposition 1. *The following statements are equivalent:*

1. P_X and Π_X are consistent, i.e. $P_X(U) \leq \Pi_X(U)$ for all $U \in \mathcal{S}(\mathcal{X})$.
2. P_X and N_X are consistent
3. $P_X(\mathcal{L}_{\Pi_X}^{\alpha^-}) \leq \alpha$ for all $\alpha \in [0, 1]$.
4. $P_X(\mathcal{L}_{\Pi_X}^{\alpha^+}) \geq 1 - \alpha$ for all $\alpha \in [0, 1]$.

Notice that consistency is not a one-to-one relationship. In particular, given a possibility measure Π_X , its credal set

$$\mathcal{C}(\Pi_X) = \{P_X : P_X(U) \leq \Pi_X(U) \quad \forall U \in \mathcal{S}(\mathcal{X})\} \quad (2)$$

contains a potentially infinite number of elements. A detailed investigation of these elements is readily accessible in [10]. One important observation is that the credal set of a possibility distribution Π_X^1 is included in that of a second one Π_X^2 , i.e. $\mathcal{C}(\Pi_X^1) \subseteq \mathcal{C}(\Pi_X^2)$, if $\pi_X^1(x) \leq \pi_X^2(x)$ for all $x \in \mathcal{X}$. Then, Π_X^1 is said to be *more specific* than Π_X^2 .

A possibility measure Π_X can be constructed to provide an outer approximation of an arbitrary set of probability distributions \mathcal{P}_X representing many common descriptions of imprecise probabilities by means of a probability-possibility transformation [9, 11, 12], e.g.

Singletons specifying a precisely defined probability distribution, i.e. $\mathcal{P}_X = \{P_X\}$,

P-boxes where the cumulative distribution function is bounded by a lower and an upper bound, F_X^- and F_X^+ , i.e. $\mathcal{P}_X = \{P_X : F_X^-(x) \leq P_X(X \leq x) \leq F_X^+(x) \quad \forall x \in \mathbb{R}\}$, refer e.g. to [13],

Intervals restricting the support of the probability distribution to lie within a given interval $[a, b]$ for real values $a < b$, i.e. $\mathcal{P}_X = \{P_X : P_X(X \in [a, b]) = 1\}$, refer e.g. to [14],

Belief measures bounding the probability distribution through degrees of belief m_X on certain sets, i.e. $\mathcal{P}_X = \{P_X : \sum_{A \subseteq U} m_X(A) \leq P_X(U) \leq \sum_{A \cap U \neq \emptyset} m_X(A) \quad \forall U \in \mathcal{S}(\mathcal{X})\}$, refer e.g. to [15],

Lower previsions specifying the minimum selling and the maximum buying price, \underline{P} and \bar{P} , for a gamble, i.e. $\mathcal{P}_X = \{P_X : \underline{P}(U) \leq P_X(U) \leq \bar{P}(U) \quad \forall U \in \mathcal{S}(\mathcal{X})\}$, refer e.g. to [16],

etc. refer e.g. to [17, 18, 19] for conceptual comparisons of the different representations of imprecise probabilities.

Therefore, it also provides a unified framework which all these alternative descriptions can be translated to in order to merge them for a combined analysis.

Yet, even though it is possible, it is usually inefficient to represent a single probability distribution by a possibility distribution. Probability distributions are the most specific representations of uncertainty, whereas possibility distributions provide a very coarse description, and much information is lost during the transformation. For instance, it is generally impossible to reconstruct the original probability distribution from a possibility distribution [11]. The same is true for p-boxes, belief measures and lower previsions. Intervals, on the other hand, are represented without loss of information.

Formally, these transformations correspond to selecting a parameterized family of nested sets $U_\beta \in \mathcal{S}(\mathcal{X})$ for $\beta \in [0, 1]$ with $U_0 = \mathcal{X}$ and $U_{\beta_1} \supseteq U_{\beta_2}$ if $\beta_1 \leq \beta_2$, and to finding a necessity distribution N_X such that $N_X(U_\beta) \leq P_X(U_\beta)$ for all $P_X \in \mathcal{P}_X$ and $\beta \in [0, 1]$ in order to ensure consistency, refer e.g. to [20]. The principle of maximum specificity, which is discussed e.g. in [21, 22], justifies choosing the tight lower bound

$$\mu(U_\beta) = \inf_{P_X \in \mathcal{P}_X} P_X(U_\beta) \quad (3)$$

as the necessity. This is sufficient for defining

$$N_X(V) = \inf_{\beta \in [0,1]: U_\beta \subseteq V} \mu(U_\beta) \quad \forall V \in \mathcal{S}(\mathcal{X}) \quad (4)$$

where the infimum of the empty set is defined to be zero.

Proposition 2. N_X is a necessity measure which is consistent with the credal set \mathcal{P}_X .

Proof. From the definition, one immediately sees that $N(\mathcal{X}) = \inf_{\beta \in [0,1]: U_\beta \subseteq \mathcal{X}} \mu(U_\beta) = \mu(U_0 = \mathcal{X}) = 1$ and $N(\emptyset) = 0$. Furthermore, let $(V_k)_{k \geq 1} \subseteq \mathcal{X}$ be a countable collections of sets, then the evaluation of the third condition yields that $N_X(\bigcap_{k \geq 1} V_k) = \inf_{\beta \in [0,1]: U_\beta \subseteq (\bigcap_{k \geq 1} V_k)} \mu(U_\beta) = \inf_{k \geq 1} \inf_{\beta \in [0,1]: U_\beta \subseteq V_k} \mu(U_\beta) = \inf_{k \geq 1} N_X(V_k)$, i.e. all conditions of a necessity measure are met. From the nestedness of the U_β it follows that $N_X(U_\beta) = \mu(U_\beta)$ for all $\beta \in [0, 1]$. Now, let $P_X \in \mathcal{P}_X$ and $V \in \mathcal{S}(\mathcal{X})$, then the definition of β^* as the maximum $\beta \in [0, 1]$ for which $U_\beta \subseteq V$ implies $N_X(V) = \mu(U_{\beta^*})$. The definition of μ yields $\mu(U_{\beta^*}) = \inf_{P'_X \in \mathcal{P}_X} P'_X(U_{\beta^*}) \leq P_X(U_{\beta^*})$, and from $U_{\beta^*} \subseteq V$ it follows that $P_X(U_{\beta^*}) \leq P_X(V)$. Hence, $N_X(V) \leq P_X(V)$ and all $P_X \in \mathcal{P}_X$ are consistent with N_X . \square

As a consequence, the sets U_β form the basis for the superlevel sets $\mathcal{L}_{\Pi_X}^{\alpha+}$ of the resulting possibility distribution. In particular, $\mathcal{L}_{\Pi_X}^{\alpha+}$ is the largest set U_β satisfying $\mu(U_\beta) \leq 1 - \alpha$. The corresponding possibility distribution $\Pi_X(V) = \sup_{\beta \in [0,1]: U_\beta \cap V = \emptyset} 1 - \mu(U_\beta)$ for all $V \subseteq \mathcal{S}(\mathcal{X})$ induces the possibility density

$$\pi_X(x) = \sup_{\beta \in [0,1]: x \notin U_\beta} 1 - \mu(U_\beta) \quad \forall x \in \mathcal{X}. \quad (5)$$

which reduces to

$$\pi_X(x) = 1 - \mu(U_{\beta(x)}) \quad \forall x \in \mathcal{X} \quad (6)$$

if an explicit representation $\beta = \beta(x)$ that provides the supremum of all $\beta \in [0, 1]$ for which $x \notin U_\beta$ is given.

Example 3 (adopted from [10]). The sets $U_\beta = [\log(\beta), -\log(\beta)]$ for $\beta \in (0, 1]$ and $U_0 = \mathbb{R}$ are a family of nested sets of \mathbb{R} . Let X be distributed with zero mean and unit variance, i.e. $X \sim P_X \in \mathcal{P}_X = \{P_X : E_{P_X}[X] = 0, V_{P_X}[X] = 1\}$. From Chebyshev's inequality it follows that $P_X(U_\beta) \geq 1 - \frac{1}{\log^2(\beta)}$ for all $P_X \in \mathcal{P}_X$. Since $\beta(x) = \exp(-|x|)$ for all $x \in \mathbb{R}$, the density function of the possibility distribution resulting from the transformation is depicted in Fig. 1 and given by

$$\pi_X(x) = \begin{cases} \frac{1}{x^2} & \text{for } |x| > 1 \quad \text{and} \\ 1 & \text{otherwise.} \end{cases} \quad (7)$$

4 Families of Nested Sets

Generally, the proposed procedure is not restricted to any family of nested sets $(U_\beta)_{\beta \in [0,1]}$. Consistency is always guaranteed by Proposition 2. However, depending on the application, some families are better suited than others for obtaining meaningful results. Below, several options for the choice of the family are presented.

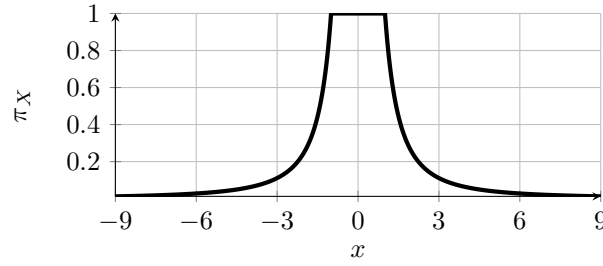


Figure 1: Possibility density function in Example 3.

4.1 General Norms

Let $\|\cdot\|$ be any norm on \mathcal{X} , e.g. the Euclidean norm, the taxicab-norm, the maximum-norm, a general p -norm, a weighted norm, combinations thereof, etc., and denote by $R > 0$ the maximum distance to a pre-defined center point $c \in \mathcal{X}$, e.g. the expected value or the origin. A suitable family is then given by $U_0 = \mathcal{X}$ and

$$U_\beta = \left\{ x \in \mathbb{R}^N : \|x - c\| \stackrel{(>)}{<} (1 - \beta) \cdot R \right\} \quad \forall \beta \in (0, 1]. \quad (8)$$

Example 4 (adopted from [11]). Let $\mathcal{P}_X = \{P_X = \mathcal{U}(c - t, c + t) : t \in (0, R]\}$ be the set of all uniform probability distributions with a symmetric support about the center point $c \in \mathbb{R}$ and the maximum distance $R > 0$ therefrom. For the family of nested sets induced by c and R , one obtains $\beta(x) = 1 - \frac{|x-c|}{R}$ for all $x \in \mathbb{R}$. The minimum probability $\mu(U_\beta) = 1 - \beta$ for all $\beta \in [0, 1]$ is obtained by considering $P_X = \mathcal{U}(c - R, c + R) \in \mathcal{P}_X$ and therefore $\pi_X(x) = 1 - \frac{|x-c|}{R}$ for $|x| \leq R$ and zero outside, i.e. a symmetric triangular possibility density on the same support and with nominal value c . Hence, the credal set of the triangular fuzzy number X with this possibility density (membership function) π_X shown in Fig. 2 contains all uniform probability distributions $\mathcal{U}(c - t, c + t)$ for $t \in (0, R]$.

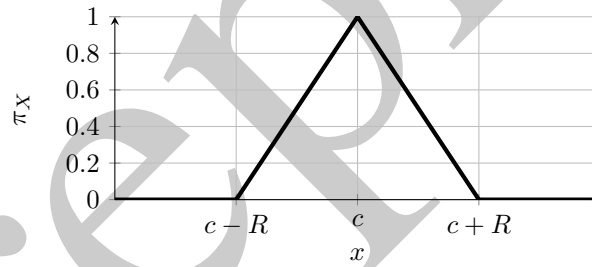


Figure 2: Triangular possibility density in Example 4.

4.2 Superlevel Sets of Probability Densities

Suppose the \mathcal{X} -valued uncertain variable X is distributed with a Borel measurable and bounded probability density function f_X , i.e. \mathcal{P}_X is a singleton. The *optimal* probability-possibility transform preserving the shape of f_X is presented in [11]. Let $\bar{f}_X = \sup_{x \in \mathcal{X}} f_X(x)$ be the maximum density of the probability density and define $U_\beta = \{x \in \mathcal{X} : f_X(x) \geq \beta \cdot \bar{f}_X\}$ for $\beta \in [0, 1]$. Then, the possibility density function $\pi_X(x) = 1 - P_X(\{\xi \in (X) : f_X(\xi) \geq f_X(x)\})$ for all $x \in \mathcal{X}$ is the optimal transformation. Dubois et al. show the following result in [11].

Proposition 5. *Let $x_1, x_2 \in \mathcal{X}$. If $f_X(x_1) \leq f_X(x_2)$, then it holds that $\pi_X(x_1) \leq \pi_X(x_2)$ for the possibility density π_X obtained by the optimal transform.*

Example 6. Let X be an \mathbb{R} -valued uncertain variable distributed with the probability distribution P_X induced by the probability density function f_X (a normalized combination of two normal distributions with means $\frac{1}{4}$ and $\frac{1}{2}$ and standard deviations $\frac{1}{12}$ and $\frac{1}{10}$, respectively). Figure 3 exhibits the possibility density function π_X obtained by the optimal transform.

4.3 Cumulative Distribution Functions

Jamison et al. argue in [23] that the cumulative distribution function of an \mathbb{R} -valued uncertain variable X is a consistent possibility density which follows immediately e.g. from $U_\beta = \left[\tanh^{-1}\left(\frac{1+\beta}{2}\right), \infty \right)$, where \tanh^{-1} is the area hyperbolic

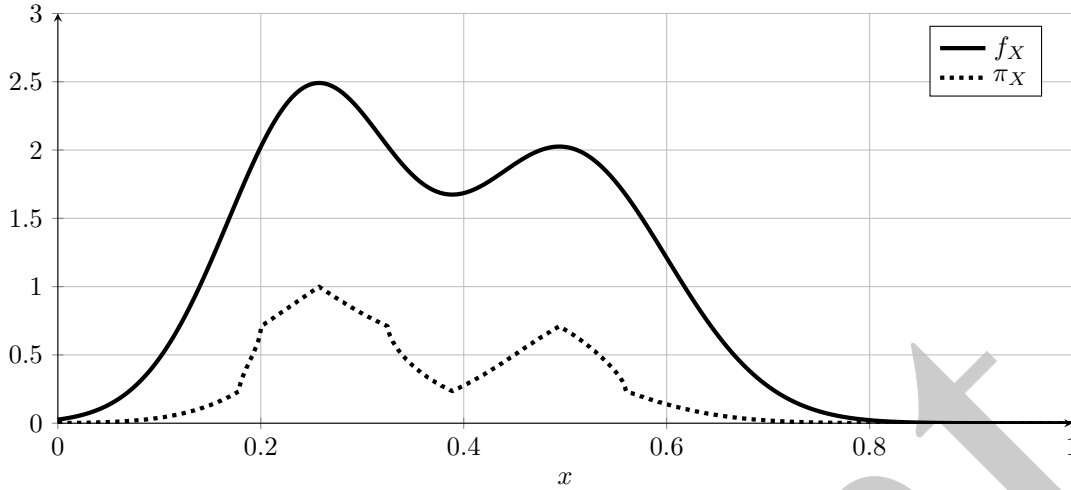


Figure 3: Probability and possibility density function in Example 6.

tangent. Then, $\beta(x) = 2 \tanh(x) - 1$ and henceforth $\pi_X(x) = \sup_{P_X \in \mathcal{P}_X} P_X(X \leq x)$ for all $x \in \mathbb{R}$. The same holds true for the complementary cumulative distribution function.

Example 7. Let $\mathcal{P}_X = \{P_X : F^-(x) \leq P_X(X \leq x) \leq F^+(x) \forall x \in \mathbb{R}\}$ be a p-box with a monotonously increasing lower and upper bound, $F_X^-, F_X^+ : \mathbb{R} \rightarrow [0, 1]$. Then, both $\pi_X^+(x) = F_X^+(x)$ and $\pi_X^-(x) = 1 - F_X^-(x)$ for $x \in \mathbb{R}$ induce consistent possibility distributions. Notice that the two distributions are not equivalent though [10].

5 Set Membership Estimation

Often, information about the distribution of an uncertain variable has to be inferred from measurements instead of employing probability-possibility transformations. Regarding the elicitation of distributions of uncertain variables, the authors argue as follows – well-aware that this opinion may be viewed as controversial:

In parameter estimation, a clear distinction between epistemic and aleatoric uncertainties has to be made, but the analysis has to commence on the basis of a purely stochastic analysis. The data-generating process itself is only subject to stochastic variability. Following the realization of a random variable therein, any imprecision about the variable's probability distribution vanishes. Consequently, data – sparse or not – exhibit only random variability and, therefore, possess a precise probability distribution. Inferring information about this distribution, is the task of the practitioner. Given data as well as a model which is supposed to explain their interplay, epistemic uncertainty is only introduced by the choice of the estimation procedure – and should always be reported. However, epistemic uncertainty is something that cannot be estimated from the data because it rather stems from the limitations inherent to the estimation process itself, such as the limited number of data. Even if the data set is big, it is not infinite, and therefore some epistemic uncertainty about the true probability distribution – as small as it may be – remains. The point of this paper is to show that possibility theory is well suited to account for both the variability due to the underlying probability distribution to be inferred and for the epistemic uncertainty introduced by the statistical estimation procedure. Before doing so in the second part of this section, some preliminary deliberations are in order.

In this section, only the direct case is considered, where the uncertain variable itself is measured. We assume that N_S realizations $(\xi_i)_{i=1}^{N_S}$ of the iid (independent and identically distributed) uncertain variables $X_1, \dots, X_{N_S} \sim P_X$ with unknown probability distribution P_X are given, $(U_\beta)_{\beta \in [0,1]}$ is a given family of nested sets and $\beta \in [0, 1]$ is fixed.

In the following, the reader may find it useful to keep in mind that the presented estimation methods basically count the number of samples that fall inside of any of these sets, and then assign a membership value to this set on the basis of probabilistic arguments about binomial distributions – a simple and intuitive idea.

5.1 Best Estimation

For a singleton credal set, the infimum in Eq. (3) vanishes, yet $\mu(U_\beta) = P_X(U_\beta)$ cannot be computed but needs to be estimated. The aim is to provide a reliable estimate of this probability by a function $\hat{\mu}$.

To answer this question, consider the $\{0, 1\}$ -valued uncertain variable $Y_\beta = \mathbb{I}_{U_\beta}(X)$, where \mathbb{I}_{U_β} is the indicator function of U_β , assuming one if $X \in U_\beta$ and zero otherwise. It is Bernoulli distributed with the success probability $P_X(U_\beta)$, and the

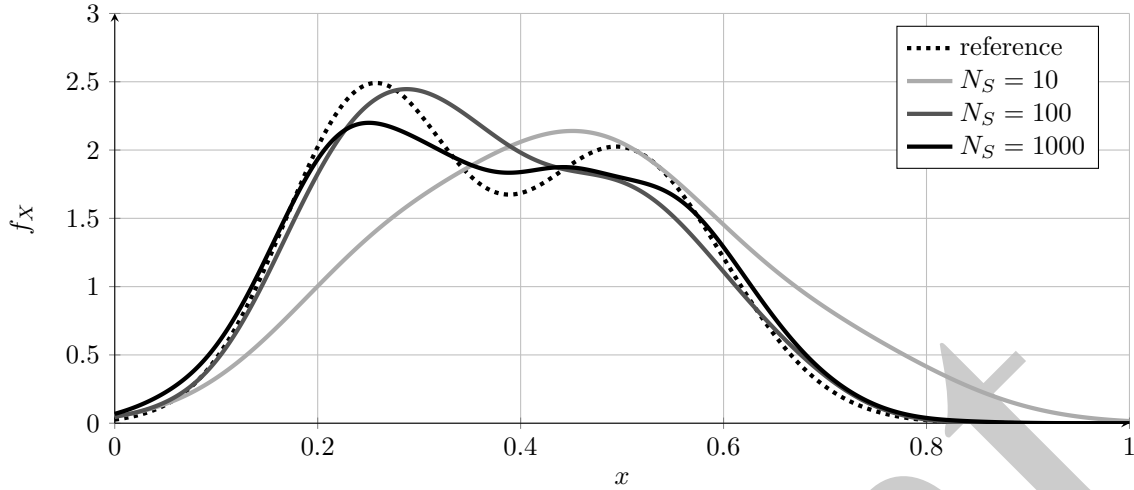


Figure 4: Reference probability density and kernel density estimates for different sample sizes N_S in Example 8.

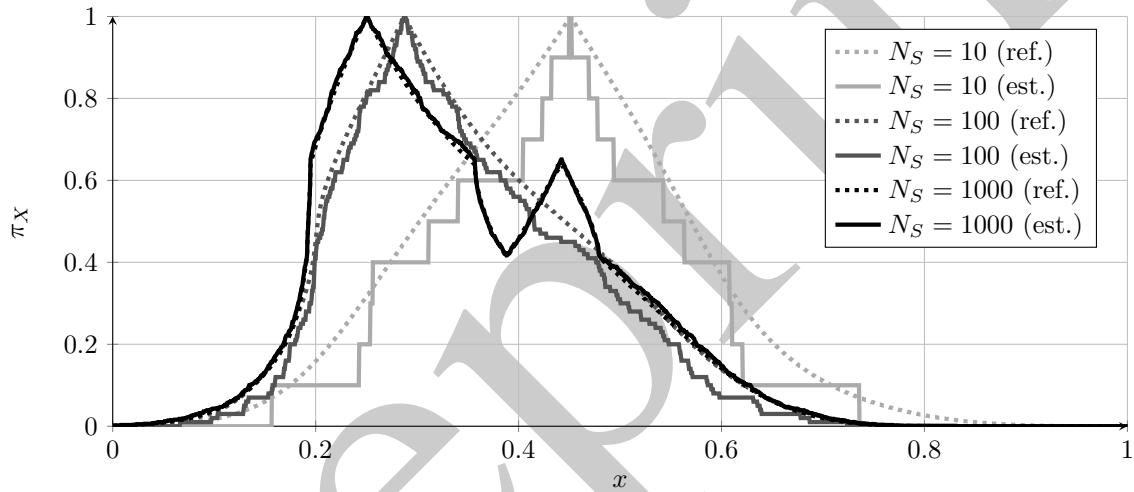


Figure 5: Reference and (best) estimated possibility densities for different sample sizes N_S in Example 8.

relative frequency of realizations with $\xi_i \in U_\beta$, i.e.

$$\hat{\mu}(U_\beta) = \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbb{I}_{U_\beta}(\xi_i), \quad (9)$$

provides an unbiased estimator of $P_X(U_\beta)$ with uniform convergence according to Borel's strong law of large numbers. Hence, approximating μ in Eq. (4) by $\hat{\mu}$ yields a possibility distribution $\hat{\Pi}_X$ which is asymptotically consistent with P_X . Refer to [24] for a more extensive discussion.

Example 8 (Continuation of Example 6). Let ξ_1, \dots, ξ_{N_S} be realizations of the \mathbb{R} -valued uncertain variables X_1, \dots, X_{N_S} distributed with the bimodal probability distribution P_X from Example 6. Since the true probability distribution P_X is assumed to be unavailable, one cannot construct the U_β precisely as explained there. However, an approximation of these sets may be obtained by replacing the probability density by a kernel density estimator \hat{f}_X with the bandwidth chosen according to Silverman's rule of thumb. The kernel density estimate is not used in the estimation of μ , only for the construction of the $U_\beta = \{x \in \mathcal{X} : \hat{f}_X(x) \geq \beta \cdot \sup_{x \in \mathbb{R}} \hat{f}_X(x)\}$ for $\beta \in [0, 1]$. Due to the non-normality of the distribution, convergence of the approximated density to the true density is rather poor as shown in Fig. 4. However, the convergence of the estimated possibility distribution to the reference possibility distribution on the respective nested set family is quite agreeable as a visual inspection of Fig. 5 suggests.

Even though possibility density estimation promises to be less involved than probability density estimation, convergence can be very poor for small numbers of N_S as Example 8 shows. This source of imprecision is neither accounted for nor indicated. Additionally, the estimation procedure presented above is a suboptimal choice since essentially one is trying to

find the possibility transform of a singleton credal set as if the relationship was indeed one-to-one. Yet, the possibility measure $\hat{\Pi}_X$ yields only the interval $P_X(U_\beta) \in [\hat{\mu}(U_\beta), 1]$ as the respective bounds for the probability. Therefore, it is unnecessary to achieve equality of $\hat{\mu}(U_\beta)$ and $P_X(U_\beta)$. Instead, one can take advantage of the whole expressiveness of possibility distributions as explained in the next subsection.

5.2 Reliable Estimation

An arguably better idea for the set membership estimation is to construct $\hat{\mu}^\gamma$ such that $\hat{\mu}^\gamma(U_\beta)$ actually serves as a reliable lower bound for $P_X(U_\beta)$, thereby accounting for the imprecision due to limited number of data. This problem is connected to finding (one-sided) confidence intervals $C^\gamma = [\hat{\mu}^\gamma(U_\beta), 1]$ for the success probability $p = P_X(U_\beta)$ of a binomial distribution given $k_\beta = \sum_{i=1}^{N_S} \mathbb{1}_{U_\beta}(\xi_i)$ successes for a sample size of $n = N_S$. The associated level of confidence $\gamma \in [0, 1]$ requires $C^\gamma = C^\gamma(k)$ to be chosen such that the coverage probability is greater than this level of confidence, i.e.

$$P_p(k \in \{0, \dots, n\} \mid p \in C^\gamma(k)) \geq \gamma \quad \forall p \in [0, 1]. \quad (10)$$

In other words, the true success probability $P_X(U_\beta)$ should be included in the confidence interval in $\gamma \cdot 100\%$ of all given scenarios for the actual number of successes. This topic is covered in most statistics text books and statistics courses, refer e.g. to [25], and a plethora of options for constructing such confidence intervals is available, refer to [26] for a comprehensive review.

Together with the normal approximation, the Pearson-Clopper interval [27] is one of the most prominent methods and, especially in the case of sparse data, it is often taken as a reference due to the guaranteed coverage probability. Many scholars, such as Agresti et al. in [28], argue – with good reason – that it can make sense to consider intervals for which the level of confidence is only approximately guaranteed. In the present setting however, the authors disagree. As mentioned above, possibility theory is a coarse description of uncertainty and should – in the authors' opinion – only be used for computing robust and reliable bounds, especially in the presence of few samples. Hence, it would be counterintuitive to employ approximations. Against this background, a one-sided variant of the Pearson-Clopper interval for a given level of confidence $\gamma \in [0, 1]$ is considered here. Its lower bound $\hat{\mu}^\gamma$ is defined through

$$B(k_\beta - 1 \mid \hat{\mu}^\gamma(U_\beta), N_S) = \gamma \quad (11)$$

for $k_\beta > 0$ and $\hat{\mu}^\gamma(U_\beta) = 0$ otherwise. Therein, the Binomial cumulative distribution function is defined as $B(k \mid p, n) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}$. The solution to Eq. (11) may be computed e.g. by means of the inverse of the beta cumulative distribution function via

$$\hat{\mu}^\gamma(U_\beta) = \begin{cases} \text{betainv}(1 - \gamma, k_\beta, N_S - k_\beta + 1) & \text{if } k_\beta > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

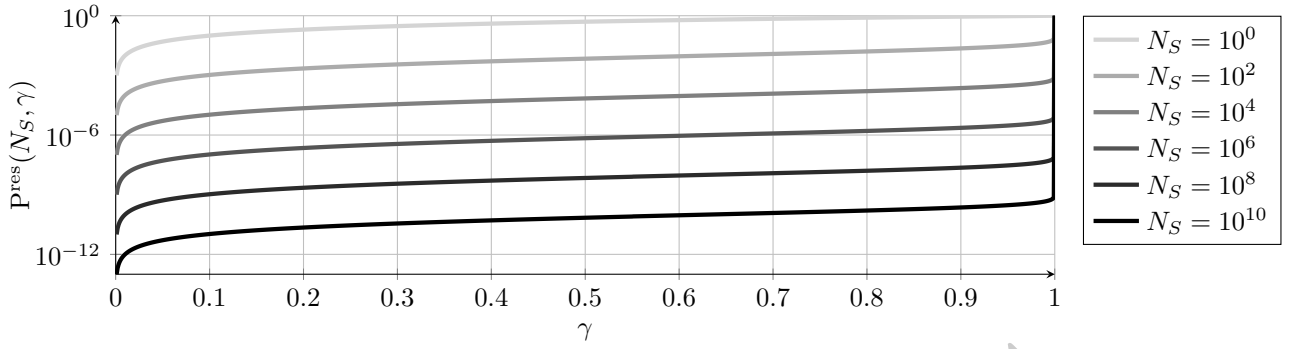
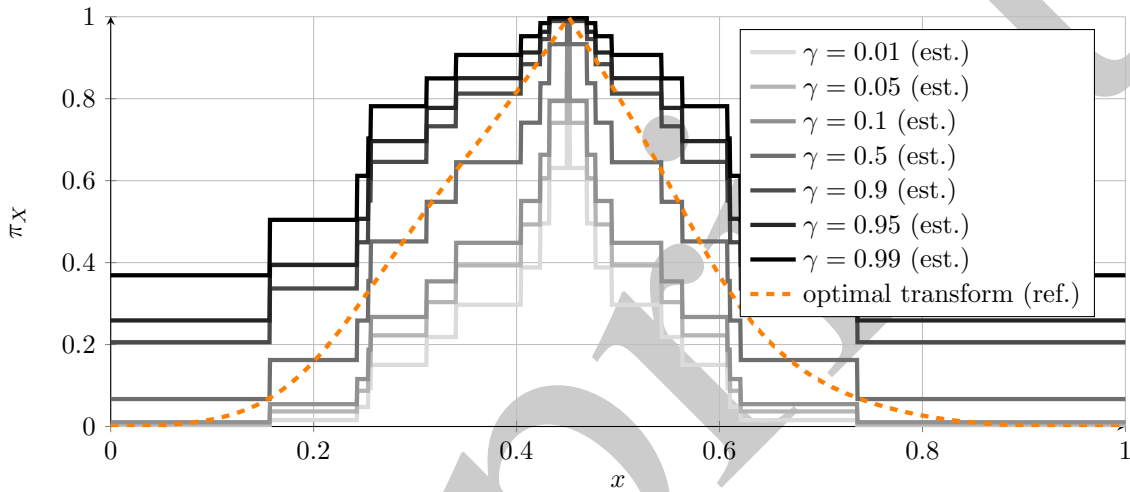
and the level of confidence γ may be viewed as the probability of consistency.

Theorem 9. *The estimated possibility distribution $\hat{\Pi}_X^\gamma$ is consistent with P_X with a guaranteed coverage probability of γ , i.e. $P_{X_1, \dots, X_{N_S}}(P_X(V) \leq \hat{\Pi}_X^\gamma(V)) \geq \gamma$ for all $V \in \mathcal{S}(\mathcal{X})$.*

Proof. The one-sided Clopper-Pearson confidence interval has a coverage probability greater than γ , see Proposition 14 in the Appendix, i.e. $P_{X_1, \dots, X_{N_S}}(\hat{N}_X^\gamma(U_\beta) \leq P_X(U_\beta)) \geq \gamma$ for all $\beta \in [0, 1]$. Let $V \in \mathcal{S}(\mathcal{X})$ and β^* be the maximum $\beta \in [0, 1]$ for which $U_\beta \subseteq V$. From the proof of Proposition 2 we know that from $\hat{N}_X^\gamma(U_{\beta^*}) \leq P_X(U_{\beta^*})$ it follows that $\hat{N}_X^\gamma(V) \leq P_X(V)$. Therefore, the probability of the latter condition being fulfilled is greater than the probability of the former condition being fulfilled, i.e. $P_{X_1, \dots, X_{N_S}}(\hat{N}_X^\gamma(V) \leq P_X(V)) \geq P_{X_1, \dots, X_{N_S}}(\hat{N}_X^\gamma(U_{\beta^*}) \leq P_X(U_{\beta^*})) \geq \gamma$, proving the theorem. \square

Of course, it is tempting to choose $\gamma = 1$. However, this will generally lead to the least specific description of uncertainty, the uniform possibility distribution on \mathcal{X} . Still, the estimated possibility distribution quickly gains in specificity by marginally lowering γ . It should also be noted that $\hat{\pi}_X^\gamma(x) > 0$ for all $x \in \mathcal{X}$. Even for U_β with $k_\beta = N_S$, i.e. which contain all samples, the lower probability $\hat{\mu}^\gamma(U_\beta)$ does not equal one – except for U_0 by assumption. There is always a chance that the next sample will lie outside due to the limited number of data at hand, which is nicely reflected by $\hat{\Pi}_X^\gamma(U) \geq P^{\text{res}}(N_S, \gamma) > 0$ for all $U \subseteq \mathcal{X}$. The actual values of this residual probability P^{res} depending on N_S and γ are depicted in Fig. 6. This should be considered when deciding on the confidence level γ .

Example 10 (Continuation of Example 8). Again, let ξ_1, \dots, ξ_{N_S} be realizations of the \mathbb{R} -valued uncertain variables X_1, \dots, X_{N_S} distributed with the bimodal probability distribution P_X from Example 6 and let U_β be the level sets of the kernel density estimates in Fig. 4. A selection of possibility densities of the reliable estimated possibility distributions $\hat{\Pi}_X^\gamma$ for different values of N_S and γ is shown in Figs. 7, 8 and 9. Notice that the possibility distributions $\hat{\Pi}_X^\gamma$ become less specific with increasing γ and, in turn, are more likely to include the reference optimal transform and, hence, are more likely to be consistent with the true probability distribution.


 Figure 6: Residual probability P^{res} in the reliable estimation procedure.

 Figure 7: Reference and (reliable) estimated possibility densities depending on γ for $N_S = 10$ in Example 10.

6 Inclusion of Measurement Models

So far, only the direct case providing measurements of the uncertain variable to be identified has been considered. Yet, in many cases only the \mathcal{Y} -valued uncertain variable $Y = Y(X)$ can be measured. Below, it is therefore assumed that only Π_Y is available, either explicitly or by estimation as $\hat{\Pi}_Y^\gamma$ or $\hat{\Pi}_Y$. Consequently, results from inverse possibilistic uncertainty propagation are necessary in order to estimate Π_X .

6.1 Precisely Specified Measurement Models

For now, suppose the probability distribution P_Y is known and the measurement function $\phi : \mathcal{X} \rightarrow \mathcal{Y}$ is a precisely known surjective and $(\mathcal{X}, \mathcal{Y})$ -measurable function. Then, there exists a (possibly infinite) number of probability distributions P_X yielding this pushforward distribution under ϕ . These extensions may be gathered in the *set of inverse probability distributions* $\mathcal{S}_{P_Y}^\phi = \{P_X : P_Y(U_Y) = P_X(\phi^{-1}(U_Y)) \forall U_Y \in \mathcal{S}(\mathcal{Y})\}$.

An analogous set of inverse possibility distributions of the possibility distribution Π_Y may also be defined. However, it is possible to account for this whole set by just one possibility distribution. The *minimum specific inverse possibility distribution* is given by

$$\Pi_X^{\text{inv.}}(U_X) = \Pi_Y(\phi(U_X)) \quad \forall U_X \in \mathcal{S}(\mathcal{X}). \quad (13)$$

In [29], it is shown that this possibility distribution possesses several properties. For instance, it is the least specific possibility distribution in the corresponding set of inverse possibility distributions. The perhaps most important result is that $\Pi_X^{\text{inv.}}$ is consistent with any probability distribution $P_X \in \mathcal{S}_{P_Y}^\phi$ from the set of inverse probability distributions of all probability distributions $P_Y \in \mathcal{C}(\Pi_Y)$ which are consistent with Π_Y .

The significance of these results for establishing a theory of inverse problems in possibility theory is e.g. discussed in [24]. In particular, therein it is argued that replacing Π_Y by $\hat{\Pi}_Y$ yields a consistent possibility distribution $\hat{\Pi}_X^{\text{inv.}}$ in the limit case for $N_S \rightarrow \infty$ since the minimum specific inverse possibility distribution preserves consistency. Analogously, replacing Π_Y by $\hat{\Pi}_Y^\gamma$ preserves the probability of consistency for $\hat{\Pi}_X^{\gamma, \text{inv.}}$.

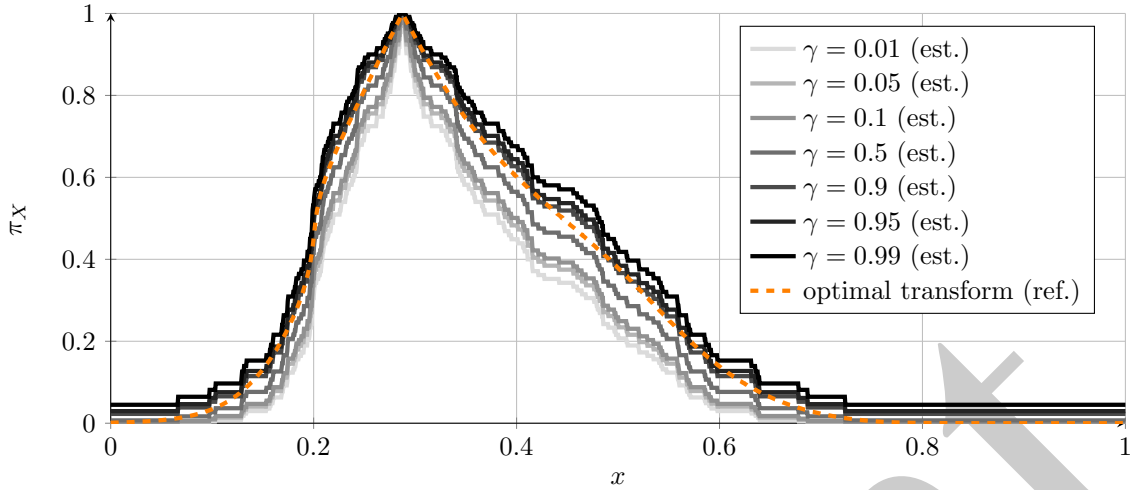


Figure 8: Reference and (reliable) estimated possibility densities depending on γ for $N_S = 100$ in Example 10.

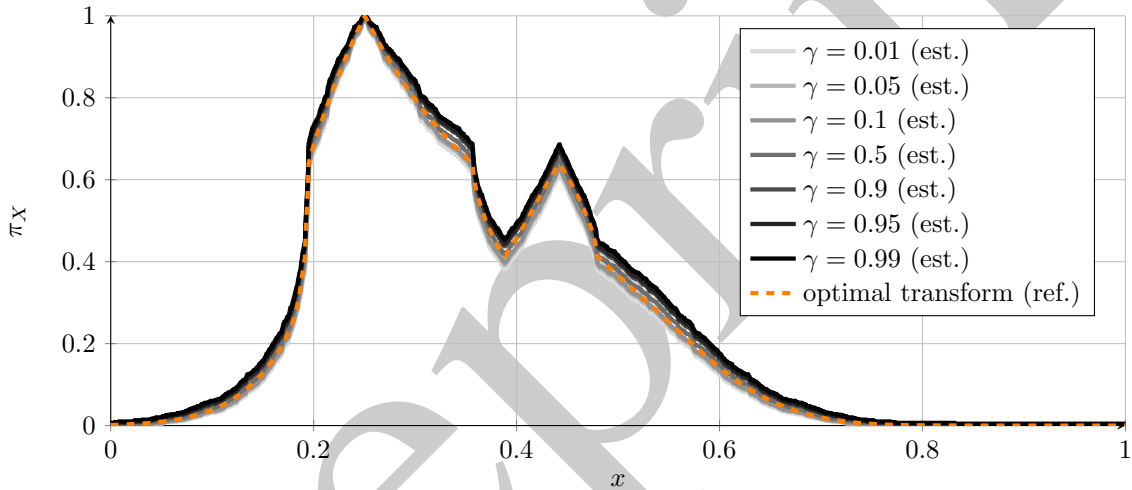


Figure 9: Reference and (reliable) estimated possibility densities depending on γ for $N_S = 1000$ in Example 10.

Example 11. Suppose $Y = X^2$ possesses the possibility density $\pi_Y(y) = 1 - y$ for $y \in [0, 1]$ and zero outside, i.e. a one-sided triangular possibility density with zero as nominal value and a radius or one to the right and zero to the left. Then, the possibility density of the minimum specific inverse possibility distribution of X is given by $\pi_X^{\text{inv.}}(x) = 1 - x^2$ for $x \in [-1, 1]$ and zero outside, see Fig. 10.

6.2 Imprecisely Specified Measurement Models

However, the solution to this precise measurement problem is of little practical relevance as the measurement function itself is often subject to uncertainty, e.g. measurement noise etc., which needs to be accounted for as well in order to establish a sensible procedure for possibilistic parameter estimation. The measurement model is then given through the $(\mathcal{X} \times \mathcal{W}, \mathcal{Y})$ -measurable function ϕ as

$$Y = \phi(X, W), \tag{14}$$

the marginal possibility distributions Π_Y and Π_W of Y and the \mathcal{W} -valued uncertain variable W are assumed to be known or to have been estimated, and the possibility distribution of X is to be inferred. Many scholars have recognized the need for a solution to this problem, refer e.g. to [30] for a review of the special case where ϕ is bilinear in X and W . Recent results in [31] facilitate the establishment of such a theory.

At best, the joint distribution $\Pi_{W,Y}$ is available. Yet, this is usually not the case. Instead, a consistent aggregation has to be performed. In [31], it is shown how probability-possibility consistency of the joint distributions may be maintained when only the marginal distributions are known. In particular, in the case of unknown copulas between the (certainly not

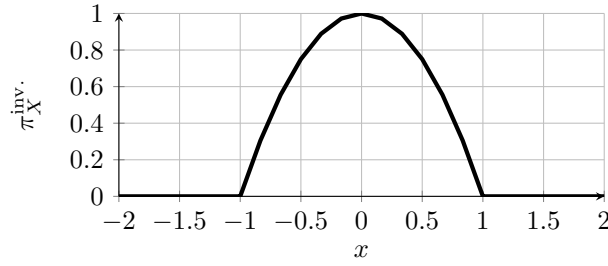


Figure 10: Possibility density of the minimum specific inverse possibility distribution of X in Example 11.

independent!) uncertain variables W and Y , the general aggregation operator yields

$$\Pi_{W,Y}(T,V) = \min(1, 2 \cdot \Pi_W(T), 2 \cdot \Pi_Y(V)) \quad \forall T \subseteq \mathcal{W}, V \subseteq \mathcal{Y}. \quad (15)$$

Therein, the correction of Zadeh's aggregation operator (the basic minimum) by the factor 2 accounts for the uncertainty dimension, i.e. the number of uncertain variables aggregated. Next, one can formally define

$$\psi: \mathcal{X} \times \mathcal{W} \rightarrow \mathcal{Y} \times \mathcal{W}, (x,w) \mapsto (\phi(x,w), w) \quad (16)$$

and compute the minimum specific inverse possibility distribution $\Pi_{X,W}^{\text{inv.}}$ under ψ . The marginalization

$$\Pi_X^{\text{inv.}}(U) = \Pi_{X,W}^{\text{inv.}}(U, \mathcal{W}) \quad \forall U \subseteq \mathcal{X} \quad (17)$$

then yields a consistent possibility distribution on X since all the above steps preserve consistency as explained above, or in [32] for the marginalization. The expression in Eq. (17) may be expressed on the basis of possibility densities as

$$\pi_X^{\text{inv.}}(x) = \sup_{w \in \mathcal{W}} \min(1, 2 \cdot \pi_W(w), 2 \cdot \pi_Y(\phi(x,w))) \quad \forall x \in \mathcal{X}. \quad (18)$$

Note that this translates loosely to Case 4 in [30], when the general aggregation operator instead of Zadeh's non-interactive aggregation operator is applied and when ϕ is bilinear in its two arguments.

In the case of simple addition with a given fuzzy-valued sum and one fuzzy-valued summand, one might be tempted to look for an equivalence to the Hukuhara difference [33] for fuzzy sets. However, this is not so straight-forward since the Hukuhara difference aims at finding the inverse of the extension principle, and the presented approach follows the paradigm of preserving consistency which, generally, leads to different results.

Example 12. Reconsider Example 11, only this time $Y = X^2 + W$ is subject to a measurement error W with a symmetric triangular possibility density $\pi_W(w) = 1 - |w|$ for $|w| \leq 1$ and zero elsewhere (center $c = 0$ and radius $r = 1$ in Fig. 2). Then the above described inference procedure yields the possibility density (see Fig. 11)

$$\pi_X^{\text{inv.}}(x) = \begin{cases} 1 & \text{if } |x| \leq 1, \\ 2 - x^2 & \text{if } 1 \leq |x| \leq \sqrt{2} \text{ and} \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

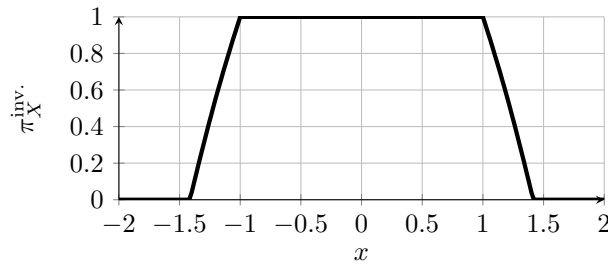


Figure 11: Possibility density of the minimum specific inverse possibility distribution of X in Example 12.

7 Computational Aspects

A key step in the estimation of the possibility distribution of the unknown input variable is the computation of the minimum specific inverse possibility distribution. A conceptually simple and elegant technique can be derived from the recursive SIVIA (Set Inversion via Interval Analysis) algorithm in [14]. An efficient implementation is described below.

Let Π_X^{inv} be the minimum specific inverse possibility distribution with respect to the output possibility distribution Π_Y and the measurement function $\phi : \mathcal{X} \rightarrow \mathcal{Y}$. The argument in the following also works in the 'imprecise measurement' case. The FSIVIA (Fuzzy Set Inversion via Interval Analysis) Algorithm 1 is – just like the SIVIA algorithm – fundamentally based on inclusion functions $[\phi]$ which provide an outer approximation of the image of intervals (or N -dimensional boxes) $[x]$ under ϕ . A detailed explanation on how to construct those would go beyond the scope of this paper and interested readers are referred to Chapter 2.4 in [14]. Here, interval arithmetic is used.

Algorithm 1: Fuzzy Set Inversion via Interval Analysis Algorithm FSIVIA.

input : Initial set $[x]$, Inclusion function $[\phi]$, Inclusion output possibility density function $[\pi_Y]$, Tolerances ε_π and ε_x

output : Subpavings $\mathcal{L}_{\Pi_X^{\text{inv}}}^-$ and $\mathcal{L}_{\Pi_X^{\text{inv}}}^+$.

```

1  $[y] \leftarrow [\phi]([x]);$ 
2  $[\pi] \leftarrow [\pi_Y]([y]);$ 
3 if  $\text{width}([\pi]) < \varepsilon_\pi$  or  $\text{width}([x]) < \varepsilon_x$  then
4    $\alpha^- \leftarrow \max([\pi]);$ 
5    $\mathcal{L}_{\Pi_X^{\text{inv}}}^- \leftarrow \{([x], \alpha^-)\};$ 
6    $\alpha^+ \leftarrow \min([\pi]);$ 
7    $\mathcal{L}_{\Pi_X^{\text{inv}}}^+ \leftarrow \{([x], \alpha^+)\};$ 
8 else
9    $([x]_1, [x]_2) \leftarrow \text{bisect } [x] \text{ where } \text{width}([x]) \text{ is maximal};$ 
10   $(\mathcal{L}_1^-, \mathcal{L}_1^+) \leftarrow \text{FSIVIA}([x]_1, [\phi], [\pi_Y], \varepsilon_\pi, \varepsilon_x);$ 
11   $(\mathcal{L}_2^-, \mathcal{L}_2^+) \leftarrow \text{FSIVIA}([x]_2, [\phi], [\pi_Y], \varepsilon_\pi, \varepsilon_x);$ 
12   $\mathcal{L}_{\Pi_X^{\text{inv}}}^- \leftarrow \mathcal{L}_1^- \cup \mathcal{L}_2^-;$ 
13   $\mathcal{L}_{\Pi_X^{\text{inv}}}^+ \leftarrow \mathcal{L}_1^+ \cup \mathcal{L}_2^+;$ 
14 end

```

Notice that the superlevel sets of Π_X^{inv} and Π_Y are linked through

$$\mathcal{L}_{\Pi_X^{\text{inv}}}^{\alpha^+} = \{x \in \mathcal{X} : \pi_X^{\text{inv}}(x) \geq \alpha\} = \{x \in \mathcal{X} : \pi_Y(\phi(x)) \geq \alpha\} = \phi^{-1}(\mathcal{L}_{\Pi_Y}^{\alpha^+}) \quad \forall \alpha \in [0, 1]. \quad (20)$$

The same holds true for the sublevel sets, i.e. $\mathcal{L}_{\Pi_X^{\text{inv}}}^{\alpha^-} = \phi^{-1}(\mathcal{L}_{\Pi_Y}^{\alpha^-})$. Therefore, one can simply apply the original SIVIA algorithm on the basis of the superlevel sets in order to obtain subpavings (partitions) $\mathcal{L}_{\Pi_X^{\text{inv}}}^-$ and $\mathcal{L}_{\Pi_X^{\text{inv}}}^+$ thereof. These subpavings are partitions of an initial interval (box) $[x] \subseteq \mathcal{X}$ where each subset is characterized by an associated α -value according to which sub- and superlevel set it belongs to. If the inclusion function of the output possibility density computes the membership values $[\pi_Y]([y]) = [\pi_Y]([\phi]([x])) = [\alpha_1, \alpha_2]$, then it is clear that $[x]$ belongs to $\mathcal{L}_{\Pi_X^{\text{inv}}}^{\alpha_2^-}$ and $([x], \alpha_2) \in \mathcal{L}_{\Pi_X^{\text{inv}}}^-$. Similarly, because all $x \in [x]$ have a membership greater than α_1 , it holds that $([x], \alpha_1) \in \mathcal{L}_{\Pi_X^{\text{inv}}}^+$. The tolerances ε_π and ε_x limit the imprecision resolution and the spatial resolution, respectively, but other termination criteria are also possible.

The sublevel subpaving may be used to reconstruct the possibility distribution via

$$\Pi_X^{\text{inv}}(U) = \sup_{([x], \alpha) \in \mathcal{L}_{\Pi_X^{\text{inv}}}^- : U \cap [x] \neq \emptyset} \alpha \quad \forall U \subseteq \mathcal{X} \quad (21)$$

and therefore it is convenient to store the subpavings in a binary tree-like structure which enables an efficient search e.g. for the marginalization in the 'imprecise measurement' case and the computation of the possibility density via

$$\pi_X^{\text{inv}}(x) = \sup_{([x], \alpha) \in \mathcal{L}_{\Pi_X^{\text{inv}}}^- : x \in [x]} \alpha \quad \forall x \in \mathcal{X}. \quad (22)$$

Example 13 (adapted from Example 3.2 in [14]). Suppose the possibility densities of two symmetric and triangular fuzzy variables with respective centers $c_1 = 10.5$ and $c_2 = 67$ and radii $R_1 = 0.5$ and $R_2 = 5$ (see Fig. 2) are given by

$$\pi_{Y_1}(y_1) = \begin{cases} 1 - \frac{|10.5 - y_1|}{0.5} & \text{if } y_1 \in [10, 11] \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \pi_{Y_2}(y_2) = \begin{cases} 1 - \frac{|67 - y_2|}{5} & \text{if } y_2 \in [62, 72] \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

They are considered non-interactive, hence Zadeh's min-aggregation

$$\pi_{y_1, y_2}(y_1, y_2) = \min(\pi_{y_1}(y_1), \pi_{y_2}(y_2)) \quad \forall y_1, y_2 \in \mathbb{R} \quad (24)$$

gives the joint possibility density. The measurement function is assumed to be

$$\phi(x_1, x_2) = \begin{pmatrix} \exp(x_1) + \exp(x_2) \\ \exp(2x_1) + \exp(2x_2) \end{pmatrix} \quad \forall x_1, x_2 \in \mathbb{R}. \quad (25)$$

The inclusion functions $[\pi_{y_1, y_2}]$ and $[\phi]$ may be obtained by the application of elementary interval arithmetic, refer to [14] or [34]. The FSIVIA algorithm is initialized on $[x] = [0, 2.5] \times [0, 2.5]$ with $\varepsilon_\pi = 0.01$ and $\varepsilon_x = 0.01$, and the subpavings of the minimum specific inverse possibility distribution shown in Fig. 12 are computed. The sublevel subpaving and the superlevel subpaving provide an upper and lower approximation of the possibility density, respectively. The projection of the sublevel set subpaving on the x_1 - and x_2 -axis, indicating the possibility densities of the marginalized possibility distributions $\Pi_{x_1}^{\text{inv.}}$ and $\Pi_{x_2}^{\text{inv.}}$, can be seen in Fig. 13.

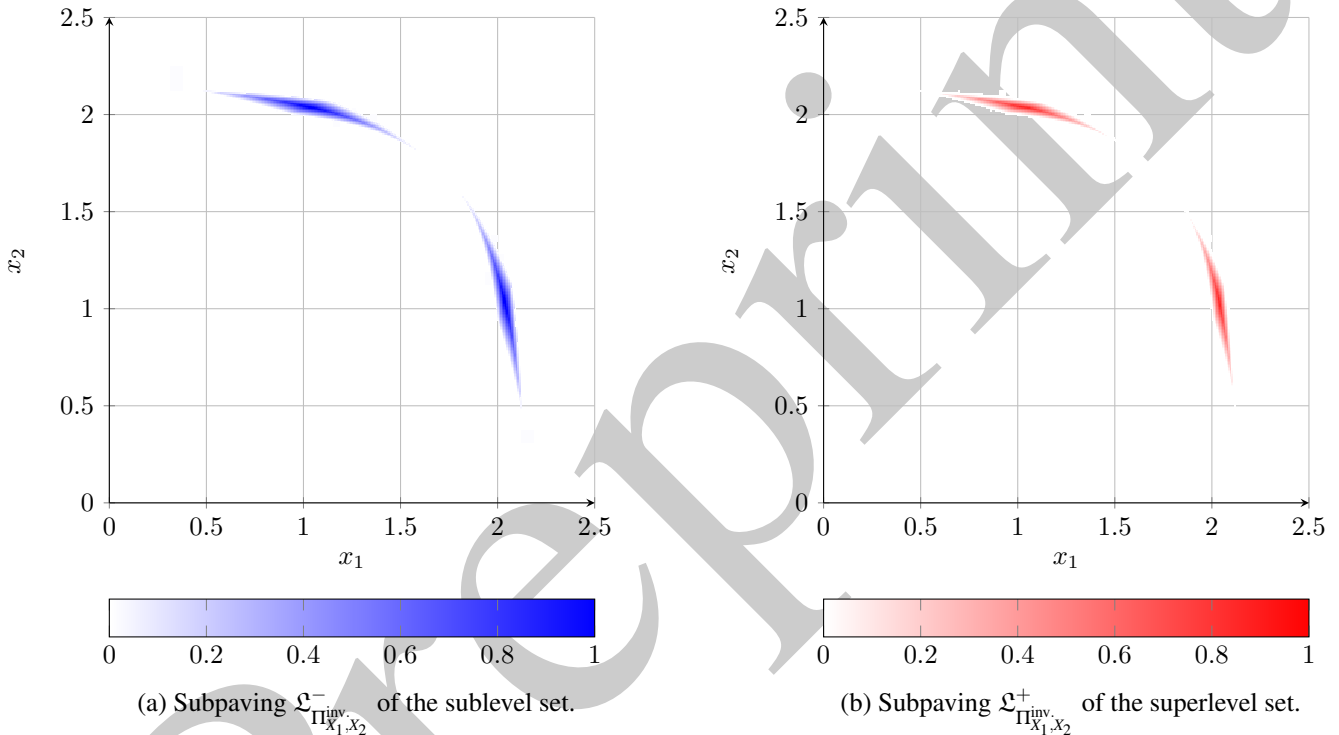


Figure 12: Level set subpavings of $\Pi_{x_1, x_2}^{\text{inv.}}$ in Example 13.

8 Possibility Distribution Estimation

Papaioannou et al. pose a benchmark problem for the quantification of polymorphic uncertainties in [35]. Therein, the probability of the failure of a wide flange steel column is investigated under aleatory as well as epistemic uncertainties, and several approaches to its solution in different frameworks are provided and compared. The problem itself is based on Example 2 in [36] and considers a column which is "subjected to a compressive load P consisting of two components $P = P_p + P_e$, where P_p denotes the permanent load and P_e the environmental load." [35]

A complete solution to the problems posed therein would go beyond the scope of this paper. However, an exemplary analysis on the basis of possibility theory with an emphasis on estimation problems shall be pursued in the following.

One uncertain variable in the mechanical model under consideration is the environmental load modeling the annual maximum snow load P_e . Its probability distribution is unknown, yet 20 measurements from consecutive years are provided to enable an estimation thereof, see Table 1. For simplicity, all approaches in [35] assume the load to follow a Gumbel distribution for which only the location and the scale parameter need to be identified. This is accomplished by several variants of Bayesian inference and maximum likelihood estimation. However, the authors of the present paper argue that 20 data points are very little to be able to assume a Gumbel distribution in good faith and one ought to pursue a non-parametric approach for the distribution estimation as described in Section 5. Further imprecision is introduced

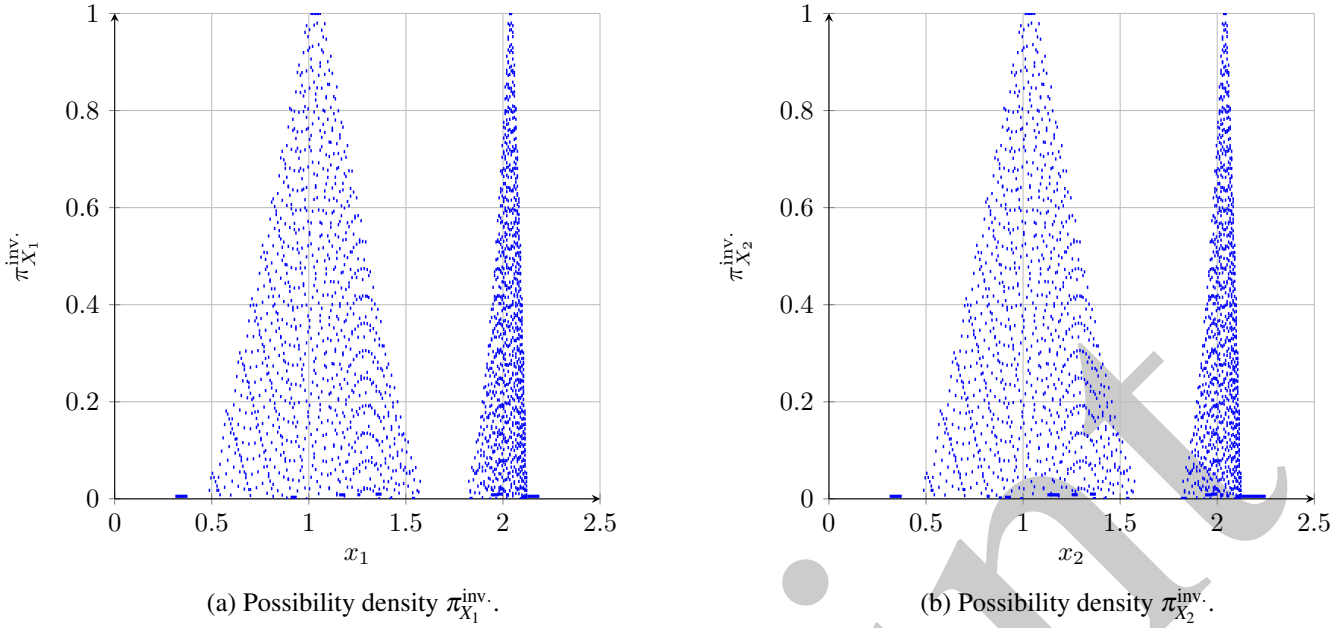


Figure 13: Marginal possibility densities obtained from projection of the sublevel set subpaving $\mathfrak{L}_{\Pi_X}^-$ in Example 13.

Table 1: Annual maximum observed snow load given in [35].

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Max load [kN]	175.6	341.3	275.2	267.6	245.2	445.5	186.1	327.7	158.3	231.7
Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Max load [kN]	150.0	269.8	263.1	246.3	272.3	202.9	191.2	191.6	245.2	176.5

by to the obvious rounding to the nearest 100 N of the data. Hence, an input variable W with an interval possibility distribution on $[-50 \text{ N}, 50 \text{ N}]$ perturbing the measured output is present. Figure 14 exhibits the corresponding reliable estimates of a consistent possibility distribution based on the 'imprecise measurement methodology' with the nested set family $U_\beta = [0, -\log(\beta)]$ for $\beta \in (0, 1)$ and $U_0 = \mathbb{R}_0^+$. For reference, also the possibility density resulting from the 'best estimation' procedure without the assumed measurement noise is given.

Above all, the estimated distributions indicate that as few as 20 samples are very little to arrive at sufficiently precise conclusions. This becomes apparent e.g. in the high residual probability $\text{Pres}(20, \cdot)$. Under this approach, the possibilistic answer to Challenge 1 in [35], posing the question if a failure probability below $1.3 \cdot 10^{-6}$ can be guaranteed, has to be a clear 'no', since the system will always fail for an environmental load which is high enough. Yet, the estimated possibility distribution for all but the lowest levels of confidence γ does not render this event improbable enough with a sufficiently low possibility.

9 Conclusions

In this paper, a possibilistic approach to data-based inference has been presented in an effort to address one of the remaining issues in possibility theory to become a general theory of polymorphic uncertainties. It is conceptually simple and easy to implement since essentially it reduces to counting how many samples fall into which of the nested sets. Again, the choice of the U_β is completely arbitrary and the main degree of freedom. But, it is a choice between 'good' and 'better', for consistency is always guaranteed.

Some remaining issues concern e.g. the identification or estimation of reference values, i.e. system parameters such as masses, stiffnesses, etc. in a possibilistic framework. The authors hope to have made a helpful contribution to the question on how such a problem could be approached. Apart from this, the presented results should also provide a meaningful approach to fuzzy-valued regression problems in the context of imprecise probabilities. Many scholars have investigated this topic, e.g. in [37, 38, 39, 40, 41, 42, 43, 44], but to the authors' knowledge none of them have pursued those by rigorously interpreting fuzzy membership functions as possibility densities of imprecisely defined probability distributions.

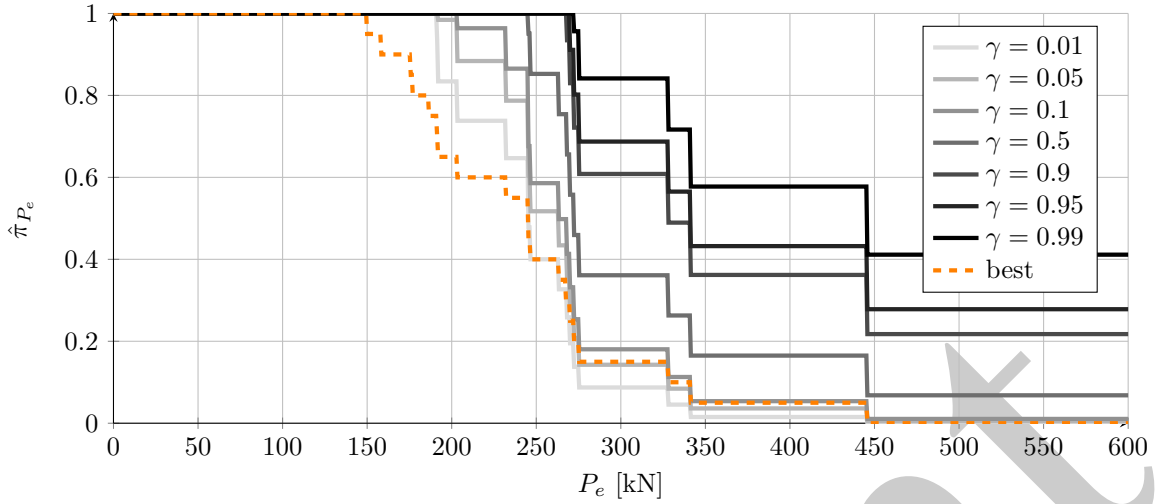


Figure 14: Reliable and best estimated possibility densities for the maximum annual snow load.

Appendix

Proposition 14 (Coverage Probability of the One-Sided Clopper-Pearson Confidence Interval). *The one-sided Clopper-Pearson confidence interval $C^\gamma(k) = [\underline{p}(k), 1]$ defined by $B(k-1 | \underline{p}(k), n) = \gamma$ for $k = 1, \dots, n$ and $\underline{p}(0) = 0$ has a coverage probability greater than γ .*

Proof. To see this, let $k \in \{1, \dots, n-1\}$. From $B(k-1 | \underline{p}(k), n) = \gamma = B(k | \underline{p}(k+1), n)$, it follows that $B(k-1 | \underline{p}(k), n) \geq B(k-1 | \underline{p}(k+1), n)$. Since B is decreasing in p , \underline{p} is monotonously increasing in k , i.e. $\underline{p}(k) \leq \underline{p}(k+1)$. Trivially, this is also fulfilled for $k = 0$. Now, let $p \in [0, 1]$ and let J be the largest j such that $p \geq \underline{p}(j)$. Hence, $\underline{p}(J) \leq p \leq \underline{p}(J+1)$. The coverage probability $P_p(k \in \{0, \dots, n\} | p \in C^\gamma(k))$ is then one for $J = n$ and otherwise it reads $\bar{P}_p(k \in \{0, \dots, J\}) = B(J | p, n)$. Once more appreciating that B is decreasing in p , it follows that $B(J | p, n) \geq B(J | \underline{p}(J+1), n) = \gamma$. \square

Acknowledgments

The authors gratefully acknowledge the support provided by the German Research Foundation (DFG) in the framework of the research project no. 319924547.

References

- [1] Thomas Haag, Jan Herrmann, and Michael Hanss. Identification procedure for epistemic uncertainties using inverse fuzzy arithmetic. *Mechanical Systems and Signal Processing*, 24(7):2021–2034, 2010.
- [2] Robert Fullér. Some properties of possibilistic linear equality and inequality systems. *Bulletin for Studies and Exchanges on Fuzziness and its Applications*, 35:80–85, 1988.
- [3] Mathieu Serrurier and Henri Prade. Maximum-likelihood principle for possibility distributions viewed as families of probabilities. In *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, pages 2987–2993. IEEE, 2011.
- [4] Hung Nguyen and Berlin Wu. *Fundamentals of Statistics with Fuzzy Data*. 2006.
- [5] Didier Dubois. Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51(1):47–69, 2006.
- [6] Marie-Hélène Masson and Thierry Denœux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157(3):319–340, 2006.
- [7] Sébastien Destercke, Didier Dubois, and Eric Chojnacki. A consonant approximation of the product of independent consonant random sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 17(06):773–792, 2009.
- [8] Didier Dubois and Henri Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49(1):65–74, 1992.

- [9] Didier Dubois and Henri Prade. Practical methods for constructing possibility distributions. *International Journal of Intelligent Systems*, 31(3):215–239, 2016.
- [10] Cédric Baudrit and Didier Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics & Data Analysis*, 51(1):86–108, 2006.
- [11] Didier Dubois, Laurent Foulloy, Gilles Mauris, and Henri Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10(4):273–297, 2004.
- [12] Didier Dubois, Henri Prade, and Sandra Sandri. On possibility/probability transformations. In *Fuzzy Logic*, pages 103–112. Springer, 1993.
- [13] Michael Beer, Scott Ferson, and Vladik Kreinovich. Imprecise probabilities in engineering analyses. *Mechanical Systems and Signal Processing*, 37(1):4 – 29, 2013.
- [14] Luc Jaulin, Michel Kieffer, Olivier Didrit, and Eric Walter. *Applied Interval Analysis: with Examples in Parameter and State Estimation, Robust Control and Robotics*, volume 1. Springer Science & Business Media, 2001.
- [15] Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pages 57–72. Springer, 2008.
- [16] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.
- [17] George Klir and Mark Wierman. *Uncertainty-based Information: Elements of Generalized Information Theory*, volume 15. Springer Science & Business Media, 1999.
- [18] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations. Part I: Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649 – 663, 2008.
- [19] Thomas Augustin, Frank PA Coolen, Gert De Cooman, and Matthias CM Troffaes. *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.
- [20] Didier Dubois and Henri Prade. Consonant approximations of belief functions. *International Journal of Approximate Reasoning*, 4(5):419 – 449, 1990.
- [21] Gilles Mauris. The principle of possibility maximum specificity as a basis for measurement uncertainty expression. In *Proceedings of IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement, AMUEM 2009*, pages 5–9. IEEE, 2009.
- [22] Didier Dubois and Henri Prade. The principle of minimum specificity as a basis for evidential reasoning. In *Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 75–84. Springer, 1986.
- [23] K David Jamison and Weldon A Lodwick. The construction of consistent possibility and necessity measures. *Fuzzy Sets and Systems*, 132(1):1–10, 2002.
- [24] Dominik Hose and Michael Hanss. Towards a general theory for data-based possibilistic parameter inference. In *Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering in Crete (Greece)*, 2019.
- [25] Philippe Rigollet. Confidence intervals for the binomial parameter p. Lecture notes in MIT mathematics course 18.650 "Fundamentals of Statistics", September 2015.
- [26] Lawrence D Brown, T Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, pages 101–117, 2001.
- [27] Charles J Clopper and Egon S Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.
- [28] Alan Agresti and Brent A Coull. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126, 1998.
- [29] Dominik Hose and Michael Hanss. Consistent inverse probability and possibility propagation. In *Proceedings of the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)*. Atlantis Press, 2019/08.
- [30] Weldon A Lodwick and Didier Dubois. Interval linear systems as a necessary step in fuzzy linear systems. *Fuzzy Sets and Systems*, 281:227–251, 2015.
- [31] Dominik Hose and Michael Hanss. Possibilistic calculus as a conservative counterpart to probabilistic calculus. *Mechanical Systems and Signal Processing*, 133:106290, 2019.
- [32] Andrey Bronevich and George J. Klir. Measures of uncertainty for imprecise probabilities: An axiomatic approach. *International Journal of Approximate Reasoning*, 51(4):365 – 390, 2010.

- 429 [33] Luciano Stefanini. A generalization of Hukuhara difference and division for interval and fuzzy arithmetic. *Fuzzy Sets*
430 *and Systems*, 161(11):1564–1584, 2010.
- 431 [34] Michael Hanss. *Applied Fuzzy Arithmetic – An Introduction with Engineering Applications*. Springer, 2005.
- 432 [35] Iason Papaioannou, Marco Daub, Martin Drieschner, Fabian Duddeck, Max Ehre, Lukas Eichner, Martin Eigel, Marco
433 Götz, Wolfgang Graf, Lars Grasedyck, et al. Assessment and design of an engineering structure with polymorphic
434 uncertainty quantification. *GAMM-Mitteilungen*, 42(2):e201900009, 2019.
- 435 [36] Norbert Kuschel and Rüdiger Rackwitz. Two basic problems in reliability-based structural optimization. *Mathematical*
436 *Methods of Operations Research*, 46(3):309–333, Oct 1997.
- 437 [37] Dominik Hose and Michael Hanss. Fuzzy linear least squares for the identification of possibilistic regression models.
438 *Fuzzy Sets and Systems*, 367:82 – 95, 2019.
- 439 [38] Wenyi Zeng, Qilei Feng, and Junhong Li. Fuzzy least absolute linear regression. *Applied Soft Computing*, 52:1009–
440 1019, 2017.
- 441 [39] Chiang Kao and Chin-Lu Chyu. A fuzzy linear regression model with better explanatory power. *Fuzzy Sets and*
442 *Systems*, 126(3):401–409, 2002.
- 443 [40] Pierpaolo D’Urso. Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. *Computational*
444 *Statistics & Data Analysis*, 42(1):47–72, 2003.
- 445 [41] Miin-Shen Yang and Hsien-Hsiung Liu. Fuzzy least-squares algorithms for interactive fuzzy linear regression models.
446 *Fuzzy Sets and Systems*, 135(2):305–316, 2003.
- 447 [42] Hideo Tanaka, Isao Hayashi, and Junzo Watada. Possibilistic linear regression analysis for fuzzy data. *European*
448 *Journal of Operational Research*, 40(3):389–396, 1989.
- 449 [43] Phil Diamond. Fuzzy least squares. *Information Sciences*, 46(3):141–157, 1988.
- 450 [44] Hisao Ishibuchi and Manabu Nii. Fuzzy regression using asymmetric fuzzy coefficients and fuzzified neural networks.
451 *Fuzzy Sets and Systems*, 119(2):273 – 290, 2001.