Constructing Consonant Predictive Beliefs from Data with Scenario Theory

Marco DeAngelis MARCO.DE-ANGELIS@LIVERPOOL.AC.UK
Institute for Risk and Uncertainty, University of Liverpool, United Kingdom

Roberto Rocchetta R.ROCCHETTA@TUE.NL
Department of Mathematics and Computer Science, Eindhoven, The Netherlands

Ander Gray ANDER.GRAY@LIVERPOOL.AC.UK
Scott Ferson SCOTT.FERSON@LIVERPOOL.AC.UK
Institute for Risk and Uncertainty, University of Liverpool, United Kingdom

Abstract

A method for constructing consonant predictive beliefs for multivariate datasets is presented. We make use of recent results in scenario theory to construct a family of enclosing sets that are associated with a predictive lower probability of new data falling in each given set. We show that the sequence of lower bounds indexed by enclosing set yields a consonant belief function. The presented method does not rely on the construction of a likelihood function, therefore possibility distributions can be obtained without the need for normalization. We present a practical example in two dimensions for the sake of visualization, to demonstrate the practical procedure of obtaining the sequence of nested sets.

Keywords: Predictive beliefs; Consonant random sets; Generalization error; Imprecise probability; Evidence theory.

1. Introduction

A very common problem in engineering science is that of computing with multivariate sample sets making the least amount of assumptions. Multivariate sample sets are difficult to deal with because even when it is possible to learn a joint probability distribution, it is often very challenging to characterize the uncertainty associated with any particular choice. We propose a method to characterize the uncertainty associated with constructing a joint distribution directly from the data using scenario theory. We show that the mechanism that constructs such joint distribution is designed to yield coherent lower predictive probabilities nicely fitting within the imprecise probability paradigm.

We present a way to construct consonant random sets–or possibility measures–that is rooted in *scenario theory* [5]. We resort to the *generalization* property of scenario theory whereby given a set of independent observations, the bounds on the predictive probability of a future observation can be computed. Recent works including [8] have elaborated on the complexity and the feasibility of such computations in highly multi-dimensional datasets. A recent overview on scenario theory can be found here [14].

This paper is also inspired by [12], whereby the problem of inferential uncertainty is cast within the imprecise probability paradigm, see also [10, 11]. Other papers on this topic are [9, 1, 13, 7]. This work focuses on two main aspects: (i) we propose a method to systematically obtain a sequence of enclosing sets with assigned probability bounds by means of scenario theory, (ii) we show that the lower predictive probability–also known as *predictive belief*–associated to the enclosing sets obtained as such, constitutes a possibility distribution, thus is a coherent lower probability. The advantage of resorting to scenario optimization to construct these sets resides in the flexibility of choosing a parametric model for the shape of the enclosing sets that are guaranteed to efficiently contain all the data. Moreover this method enables the encoding of both aleatory and epistemic dependence within the same framework.

2. Problem Statement

Let X be a random variable on a domain $\mathscr{X}$ with unknown probability distribution $\mathbb{P}_X$. We would like to quantify the beliefs held by an agent about a random future realization of X from past observations $X_1,...,X_n$ randomly drawn from the same distribution and mutually independent (*iid*). This question is typically addressed within probability theory either using Bayesian inference or with classical statistical inference, however when the sample size is limited neither approach yields a unique answer, without additional modelling assumptions, especially in the multivariate case. With scenario theory we aim to characterize the uncertainty about the distribution $\mathbb{P}_X$ using a consonant predictive belief function given some random data (scenarios).

2.1. Predictive Beliefs and Belief Functions

Given *iid* random samples $X_1,...,X_n$, with parent probability distribution $\mathbb{P}_X$, we want to produce a belief function, denoted by $\mathsf{Bel}_X := \mathsf{Bel}(\cdot; X_1, \cdots, X_n)$, in such a way that the inequality $\mathsf{Bel}_X \leq \mathbb{P}_X$, will hold at least in $100(1-\beta)\%$ of the cases. Which properties should Bel_X satisfy? As more and more data are gathered, it seems reasonable to postulate that $\mathsf{Bel}_X = \mathbb{P}_X$, from a frequentist perspective [7]. A sample set of infinite size should be equivalent to knowledge of the distribution of X, hence the belief function should asymptotically become identical to $\mathbb{P}_X$. For finite n, it seems natural to impose that Bel_X will be less committed than $\mathbb{P}_X$, which leads to $\mathsf{Bel}_X \leq \mathbb{P}_X$. However, because of limited knowledge, the inequality is often too stringent to hold in all cases. In other terms, assuming the experiment is repeated indefinitely, we demand Bel_X to be less committed than $\mathbb{P}_X$ *most of the time*, with at least some prescribed long run frequency $1-\beta$, where $\beta \in (0,1)$ is a small number. In summary, the following two requirements are typically imposed to the belief function: (1) $\forall A \subseteq \mathscr{X}, \mathsf{Bel}_X(A) \to \mathbb{P}_X(A)$, as $n \to \infty$, and (2) $\mathbb{P}^n(\mathsf{Bel}_X \leq \mathbb{P}_X) \geq 1-\beta$. A belief function satisfying these two requirements is called a predictive belief function at confidence level $1-\beta$. Several methods have been proposed to construct these belief functions, like in [7], using multinomial confidence intervals, and then further extended to the continuous case [1]. In this work, we restrict our construction to sets that are mutually nested, and resort to scenario theory to compute the belief measure. The consonant belief function obtained as such will therefore satisfy the second requirement, but not the first requirement. Let us recall now that belief functions are coherent lower probabilities. Given a power set $2^{\mathscr{X}}$, the function $m: 2^{\mathscr{X}} \to [0,1]$ is called a *basic belief assignment*, if the following two properties hold: (i) $m(\emptyset) = 0$, (ii) $\sum_{A \in 2^{\mathscr{X}}} m(A) = 1$. The mass $m(A)$ can be interpreted as the part of the agent's belief allocated exactly to the hypothesis that X takes some value in A [15]. The subsets $A \subseteq \mathscr{X}$ such that $m(A) > 0$ are called the *focal sets* of m. For any focal set $A \in 2^{\mathscr{X}}$, its belief can be obtained summing up the masses of the focal elements $B \in 2^{\mathscr{X}}$ contained in it,

$$\mathsf{Bel}_X(A) = \sum_{B: B \subseteq A} m(B). \tag{1}$$

When the focal sets are nested, Bel_X is said to be *consonant*. Consonance has recently been used to build confidence preserving structures [3]. Eq. 1 constitutes a lower bound on the true yet unknown probability $\mathbb{P}_X(A)$ to the hypothesis that X takes some value in A. Thus the belief function in Eq. 1 falls in a subclass of coherent lower probabilities [2, 6]. We propose an algorithm to construct recursively enclosing sets of predictive beliefs, which are also known as *necessities*, directly from data. The resulting consonant structure provides a lower bound on the probability of observing new data in any given enclosing set. Note that by limiting this approach to consonant beliefs, the second reequirement cannot be reached for all events, even when $n \to \infty$. This is because the consonant belief structure is effectively a possibility distribution, which always corresponds to a multitude of probability distributions, thus never converges to a single probability distribution [16].

3. Scenario Theory

Consider a closed bounded set $B \subseteq \mathbb{R}^m$, like one of the rectangles in Figure 1 where $m = 2$. Let us consider a vector of parameters $z \in \mathscr{Z} \subseteq \mathbb{R}^d$ that uniquely identify B, e.g. its size and location. Given a convex cost function $f(z)$ e.g. the area of the rectangle, a family of convex constraints $z \in \mathscr{Z}_X \subseteq \mathbb{R}^d$, parametrized by the data X, and a sample set $X_1, ..., X_n$ with $d < n$, of *iid* realizations, the following optimization program

$$\min_{z \in \mathscr{Z}} f(z) \quad \text{subject to: } z \in \bigcap_{i=1,\dots,n} \mathscr{Z}_{X_i}, \tag{2}$$

determines the optimal set B, while $X_1, ..., X_n$ are called *scenarios*, which have to be interpreted as observations from which one wants to make a decision, i.e. select an optimal value of z. Eq. 2 is a standard convex problem that can be numerically solved with common software. From this point on, the optimization of Eq.2 will only be used to identify the number of scenarios k that are *active* constraints, i.e. those points that prevent the cost function from further improving.

Definition 1 (Enclosing set of degree k) *The optimal set $B \subseteq \mathbb{R}^m$, that strictly contains $n-k$ observations is referred to as an enclosing set of degree k, and it is denoted by B_k, where k is the number of observations strictly not contained in B_k, i.e. either on the borders or outside B_k.*

The observations lying exactly on the border of B_k are also known as *active scenarios*. Note that k is a random variable because it is a function of the data, while B_k is a realization of the random set expressed by the belief function. Given a parametric set B, e.g. a hyper-box or an ellipsoid, the optimization program in Eq.2 is tasked to identify the number of active scenarios for a given subset of observations that are to be enclosed. If the optimization problem is convex the optimal set is unique and can be computed exactly.

Definition 2 (Lower bound probability of degree k) *The precise predictive probability of a given enclosing set of degree k: $\mathbb{P}_X(B_k)$, has a lower bound $\underline{p}_k$, with assigned one-sided coverage probability. For a fixed and arbitrarily small $\beta \in (0,1)$, and a number $k \in \mathbb{Z}_+$ of active scenarios the coverage probability of the given set B_k is*

$$\mathbb{P}^n\left(\underline{p}_k \leq \mathbb{P}_X(B_k)\right) \geq 1-\beta, \tag{3}$$

Where $\mathbb{P}^n$ is the product probability $(X_1, ..., X_n)$ due to independence of the samples. The (one-sided) lower bound $\underline{p}_k$ is a random variable of the probability space of the data generating mechanism and can be computed using the formula and code provided in [4]. A code for the two-sided case for both lower and upper bounds is provided in [8]. The solution t_k to the polynomial equation,

$$\frac{\beta}{n+1}\sum_{m=k}^{n}\binom{m}{k}t^{m-k}-\binom{n}{k}t^{n-k}=0, \tag{4}$$

is unique in the interval $(0,1)$. Under the two assumptions of *uniqueness* and *non-degeneracy*, it holds that $p_k = t_k$, see Theorem 2 in [4]. To compute t_k a simple bisection algorithm can be used. More about the two assumptions can also be found in [4]. The positive integer k can be interpreted in multiple ways: (i) as a parameter defining how complex is our decision z; (ii) the minimum number of scenarios/samples that are necessary to reconstruct the optimized z; (iii) an over-fitting parameter which constitutes the key for constructing consonant sets as described in the next section.

4. Method for Constructing Consonant Predictive Beliefs

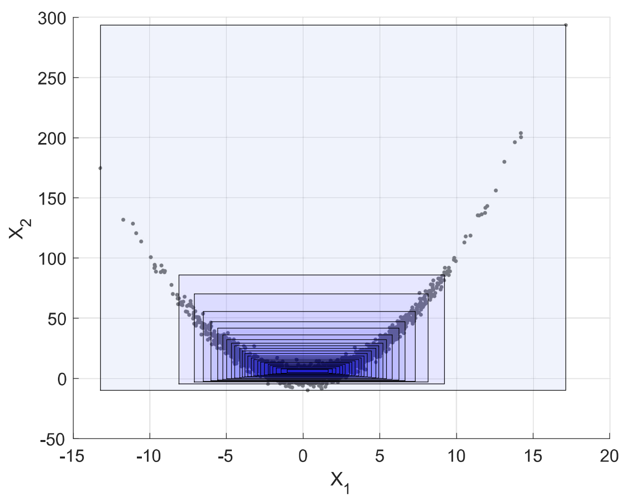

Figure 1: Consonant box-shaped enclosing sets.

The proposed method for constructing sequentially enclosing sets is very straightforward. Let us consider the observations $X_i \in \mathbb{R}^2$ with $i = 1, \cdots, n$ shown as scatter plot in Figure 1. A simple shape for the enclosing set is for example a rectangle. The rectangle with the smallest area enclosing the dataset X_i is uniquely determined by at most k observations, which are the so called *active scenarios*. We denote this rectangle by B_k. Note that in the case of a rectangle the number of support scenarios is at least two. Also note that the value of k is data dependent (random) and is affected by the functional form of B. Given a fixed and arbitrarily small $\beta \in (0,1)$ the predictive belief $\underline{p}_k$ associated with B_k can be computed finding the roots of the polynomial in Eq.3, which is independent on the shape of the set. the base of the Aztec pyramid shown from above in Figure 1 corresponds to the rectangle-shaped set enclosing all the observations. This set is associated with the largest predictive belief. The algorithm that constructs the sequence of nested rectangles proceeds by cumulating the number of active scenarios k until all the observations have been converted to support scenarios. In particular we propose to use a simple arithmetic progression with common difference equal to the number of active scenarios at the given iteration, to encourage symmetry in the resulting sequence of nested sets. The second iteration determines a rectangle whose support scenarios are sought in the dataset stripped of the observations that were identified as active scenarios at the previous iteration. This procedure is repeated until no more observations can be found in a given set, thus guaranteeing consonance of the constructed rectangles. In Figure 2 the two marginal plausibility distributions that result from the procedure are shown. Because of Eq.3, any sequence of active scenarios verifying $0 = k_0 < k_1 < \cdots < k_n = n$, implies

$$1 \geq \underline{p}_{k_0} \geq \underline{p}_{k_1} \geq \cdots \geq \underline{p}_{k_n} = 0, \tag{5}$$

$$\mathscr{X} \supseteq \underline{B}_{k_0} \supseteq \underline{B}_{k_1} \supseteq \cdots \supseteq \underline{B}_{k_n} = \emptyset. \tag{6}$$

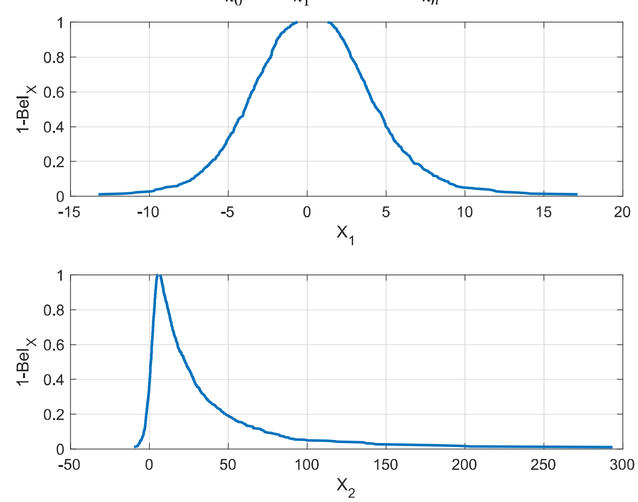

Figure 2: Plausibility contours obtained from Eqs.5 and 6, projecting onto the coordinate axes.

4.1. The Lower Bound $\underline{p}_k$ Is a Predictive Belief

We are ready to state the main result of this paper:

Theorem 3 *The lower bound $\underline{p}_k$ is a sequence of valid predictive beliefs for any $k \in \mathbb{Z}_+$ such that $0 = k_0 < k_1 < \cdots < k_n = n$.*

Proof Eq.3 guarantees that the lower bound $\underline{p}_k$ is always non-increasing with k, see also Eq. 5. Two sets obtained from the same scenario optimization program, but with different number of active scenarios are always weakly nested to one another , see Eq. 6. This suffices to provide a strong condition for the sequence of lower bounds $\underline{p}_k$ to be valid beliefs. ■

Acknowledgements

The authors would like to thank Dominik Hose and Alexander Wimbush for the fruitful discussions on possibility distributions and likelihood functions; and Ryan Martin for the insightful review of the manuscript. The work was supported by the "ITEA3-2018-17030-DayTiMe" project and by the UKRI Engineering & Physical Sciences Research Council with grant no. EP/R006768/1.

References

[1] A. Aregui and T. Denœux. Constructing predictive belief functions from continuous sample data using confidence bands. In *5th International Symposium on Imprecise Probability: Theories and Applications*, 2007.

[2] T. Augustin, F.P.A. Coolen, G. De Cooman, and M.C.M. Troffaes. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

[3] M. S. Balch. Mathematical foundations for a theory of confidence structures. *International Journal of Approximate Reasoning*, 53(7):1003–1019, 2012.

[4] M. C. Campi and S. Garatti. Wait-and-judge scenario optimization. *Mathematical Programming*, 167(1): 155–189, Jan 2018. ISSN 1436-4646. doi: 10.1007/s10107-016-1056-9.

[5] M. C. Campi, S. Garatti, and F. A. Ramponi. A general scenario theory for nonconvex optimization and decision making. *IEEE Transactions on Automatic Control*, 63(12):4067–4078, Dec 2018. ISSN 0018-9286. doi: 10.1109/TAC.2018.2808446.

[6] T. Denœux. Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning*, 20(1):79–111, 1999. ISSN 0888-613X. doi: https://doi.org/10.1016/S0888-613X(00)88944-6.

[7] T. Denœux. Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, 42 (3):228–252, 2006. ISSN 0888-613X. doi: https://doi.org/10.1016/j.ijar.2006.01.001.

[8] S. Garatti and M. C. Campi. Risk and complexity in scenario optimization. *Mathematical Programming*, pages 1–37, November 2019. ISSN 1436-4646. doi: 10.1007/s10107-019-01446-4.

[9] D. Hose and M. Hanss. On data-based estimation of possibility distributions. *Fuzzy Sets and Systems*, 399: 77–94, 2020. ISSN 0165-0114. doi: https://doi.org/10.1016/j.fss.2020.03.017. Fuzzy Intervals.

[10] C. Liu and Ryan M. Inferential models and possibility measures, 2020.

[11] R. Martin. An imprecise-probabilistic characterization of frequentist statistical inference, 2021.

[12] R. Martin and C. Liu. *Inferential Models: Reasoning with Uncertainty*. Chapman Hall/CRC Monographs on Statistics and Applied Probability. CRC Press, 2015. ISBN 9781439886519.

[13] M.H. Masson and T. Denœux. Inferring a possibility distribution from empirical data. *Fuzzy sets and systems*, 157(3):319–340, 2006.

[14] R. Rocchetta, Q. Gao, and M. Petkovic. Soft-constrained interval predictor models and epistemic reliability intervals: A new tool for uncertainty quantification with limited experimental data. *Mechanical Systems and Signal Processing*, 161:107973, 2021. ISSN 0888-3270. doi: https://doi.org/10.1016/j.ymssp.2021.107973.

[15] G. Shafer. *A Mathematical Theory of Evidence*. Limited paperback editions. Princeton University Press, 1976. ISBN 9780691100425.

[16] M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52 (6):767–791, 2011. ISSN 0888-613X. doi: https://doi.org/10.1016/j.ijar.2011.02.001.