



Multistep Prediction using Point-Cloud Approximation of Continuous Belief Functions

Thierry Denoeux, Orakanya Kanjanatarakul

► To cite this version:

Thierry Denoeux, Orakanya Kanjanatarakul. Multistep Prediction using Point-Cloud Approximation of Continuous Belief Functions. IEEE International Conference on Fuzzy Systems (FUZZ 2019), Jun 2019, New Orleans, United States. pp.1-6, 10.1109/FUZZ-IEEE.2019.8858988 . hal-02471564

HAL Id: hal-02471564

<https://hal.science/hal-02471564>

Submitted on 5 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multistep Prediction using Point-Cloud Approximation of Continuous Belief Functions

Thierry Denoeux

Université de Technologie de Compiègne, CNRS
UMR 7253 Heudiasyc, Compiègne, France
Email: tdenoeux@utc.fr

Orakanya Kanjanatarakul

Chiang Mai Rajabhat University
Chiang Mai, Thailand
Email: orakanyaa@gmail.com

Abstract—We consider the problem of quantifying prediction uncertainty in the Dempster-Shafer framework. Our approach assumes a parametric statistical model relating the variable of interest, the parameter and a pivotal random variable with known probability distribution. A predictive belief function is computed using this model and a belief function defined in the parameter space. In the case of multistep prediction, the quantity to be predicted is a vector, and the predictive belief function is defined in a multidimensional space, making its representation and manipulation difficult. To address this issue, we propose to approximate the focal sets of belief functions using point clouds, which allows us to approximate the belief and plausibility of arbitrary events with any accuracy. As an illustration, the approach is applied to the case of a first-order autoregressive process.

Index Terms—Dempster-Shafer theory, evidence theory, imprecise probabilities, possibility measure, forecasting, point clouds, uncertainty quantification, autoregressive process.

I. INTRODUCTION

In many areas, it is important to guess the values of some not yet observed variables in order to support decision-making. A good prediction system not only determines the quantities of interest with low error most of the time, but it is also capable of *quantifying the uncertainty* of the predictions. In previous work [1], [2], we have argued for the Dempster-Shafer theory of belief functions [3], [4] as a suitable framework for quantifying prediction uncertainty. The approach introduced in [1] postulates the existence of a parametric statistical model describing the data generating process. Such a model typically allows us to describe the variable Y of interest as a function of some parameter θ and a random variable U with known distribution. For instance, if Y has a normal distribution with mean μ and standard deviation σ , it can be written as

$$Y = \mu + \sigma\Phi^{-1}(U) = \varphi(\theta, U), \quad (1)$$

where $\theta = (\mu, \sigma)$, Φ is the standard normal cumulative distribution function (cdf) and U has a standard uniform distribution in $[0, 1]$. An equation such as (1) will be referred to as a *φ -equation*. Typically, parameter θ is unknown, but a belief function on θ can be constructed from past observations using the likelihood function [4], [5]. This belief function, together

with the probability distribution of U , can be propagated through the φ -equation to compute a belief function Bel_Y on Y .

In [1] and [2], we have described numerical schemes allowing us to approximate Bel_Y through Monte Carlo simulation and constrained optimization. These schemes, however, can only be used when Y is a scalar quantity and the focal sets of Bel_Y are real intervals. In some problems such as multistep time series forecasting, Y is an h -dimensional random vector, with $h > 1$. The belief function Bel_Y is then still well defined mathematically, but its focal sets are regions of \mathbb{R}^h with arbitrary shape. Computing the degree of belief $Bel_Y(B)$ of events B of interest then becomes more challenging. In this paper, we propose to tackle this problem by combining Monte Carlo simulation and a representation of focal sets using *point clouds*.

The rest of this paper is organized as follows. The necessary background on belief functions will first be recalled in Section II. The application to prediction and the proposed scheme for approximating a multidimensional predictive belief function will then be described in Section III. Finally, as an illustration, our approach will be applied to multistep time series prediction with an $AR(1)$ model in Section IV, and concluding remarks will be given in Section V.

II. BELIEF FUNCTIONS

To make the paper self-contained, the basic notions of belief and plausibility functions will first be recalled in Section II-A. The relation with random sets, which is instrumental in our approach, will then be explained in Section II-B, and the special case of consonant belief functions will be addressed in Section II-C. Finally, Dempster's rule will be recalled in Section II-D.

A. Definitions

Let Ω be a set (not necessarily finite) and \mathcal{B} an algebra of subsets of Ω (i.e., a nonempty family of subsets of Ω , closed under complementation and finite intersection). A *belief function* on \mathcal{B} is a mapping $Bel : \mathcal{B} \rightarrow [0, 1]$ verifying $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$ and the complete monotonicity

property: for any $k \geq 2$ and any collection B_1, \dots, B_k of elements of \mathcal{B} ,

$$Bel\left(\bigcup_{i=1}^k B_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} B_i\right).$$

A function $Pl : \mathcal{B} \rightarrow [0, 1]$ is a *plausibility function* iff the mapping $B \rightarrow 1 - Pl(\bar{B})$ is a belief function. In Dempster-Shafer theory [4], belief functions are used to represent an agent's belief about some unknown quantity Y taking values in Ω , based on some evidence. The number $Bel(B)$ for any event $B \in \mathcal{B}$ is interpreted as the total degree of support in B , while $Pl(B)$ measures the lack of support in the complement of B . Complete ignorance is, thus, represented by the *vacuous* belief function verifying $Bel(B) = 0$ for all $B \neq \Omega$, and $Pl(B) = 1$ for all $B \neq \emptyset$.

B. Random set representation

A belief function can always be induced by a *random set* [3], [6], [7]. This point of view is particularly useful to define and manipulate belief functions on continuous spaces such as \mathbb{R}^p [2]. More precisely, let S be a state space, \mathcal{A} an algebra of subsets of S , and \mathbb{P} a finitely additive probability on (S, \mathcal{A}) . A mapping Γ from S to 2^Ω (referred to as a *multivalued mapping*) is said to be *strongly measurable* with respect to (S, \mathcal{A}) and (Ω, \mathcal{B}) if, for any $B \in \mathcal{B}$, the set

$$\Gamma_*(B) = \{s \in S \mid \Gamma(s) \neq \emptyset, \Gamma(s) \subseteq B\}$$

belongs to \mathcal{A} . Then, the mapping \mathbb{P}_* from \mathcal{B} to $[0, 1]$ defined as

$$\mathbb{P}_*(B) = \frac{\mathbb{P}(\Gamma_*(B))}{\mathbb{P}(\Gamma_*(\Omega))}, \quad (2)$$

for all $B \in \mathcal{B}$, is a belief function [6]. The sets $\Gamma(s)$ are called the *focal sets* of \mathbb{P}_* .

C. Consonant belief functions

The random-set view of belief functions outlined in the previous section provides a way to define practical models using well-chosen probability spaces $(S, \mathcal{A}, \mathbb{P})$ and multivalued mappings Γ . In particular, let S be the interval $[0, 1]$, $\mathcal{A} = \beta_{[0,1]}$ the Borel σ -field on S , $\mathbb{P} = \lambda$ the Lebesgue measure on $[0, 1]$, and π a mapping from Ω to S s.t. $\sup_{\omega \in \Omega} \pi(\omega) = 1$. Then, the mapping $\Gamma : S \mapsto 2^\Omega$ defined by

$$\Gamma(s) = \{\omega \in \Omega \mid \pi(\omega) \geq s\}$$

for all $s \in [0, 1]$ is strongly measurable [8]. In this case, the induced belief function is consonant, i.e., for any s and s' in $[0, 1]$, we have either $\Gamma(s) \subseteq \Gamma(s')$ or $\Gamma(s') \subseteq \Gamma(s)$, and the corresponding plausibility function is a *possibility measure* [9]; it can be computed as $Pl(B) = \sup_{\omega \in B} \pi(\omega)$ for all $B \subseteq \Omega$. In particular, $Pl(\{\theta\}) = \pi(\theta)$. (Function π is a possibility distribution; it is often referred to in this context as the *contour function*). This model is useful, for instance, in statistical inference, where π can be taken to be the relative likelihood function [4], [5], yielding a consonant belief function on the parameter space.

D. Dempster's rule

Dempster-Shafer theory is fundamentally based on the idea of combining belief functions induced by independent pieces of evidence [3], [4], [10]. Given two belief functions Bel_1 and Bel_2 on Ω generated by random sets $(S_i, \mathcal{A}_i, \mathbb{P}_i, \Gamma_i)$, $i = 1, 2$, the *orthogonal sum* of Bel_1 and Bel_2 , denoted by $Bel_1 \oplus Bel_2$, is the belief function generated by the random set $(S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2, \Gamma_\cap)$, where $\mathcal{A}_1 \otimes \mathcal{A}_2$ is the algebra generated by sets $A_1 \times A_2$, with $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, $\mathbb{P}_1 \otimes \mathbb{P}_2$ is the product of \mathbb{P}_1 and \mathbb{P}_2 , and

$$\Gamma_\cap(s_1, s_2) = \Gamma_1(s_1) \cap \Gamma_2(s_2).$$

Practically, Dempster's rule often needs to be approximated by Monte Carlo simulation [2], [11]. Typically, we draw pairs (s_1, s_2) from $\mathbb{P}_1 \otimes \mathbb{P}_2$ and keep them only if $\Gamma_\cap(s_1, s_2) \neq \emptyset$. After obtaining N such pairs $(s_1^{(1)}, s_2^{(1)}), \dots, (s_1^{(N)}, s_2^{(N)})$, we get an approximate representation of $Bel_1 \oplus Bel_2$ as N non-empty focal sets $\Gamma_\cap(s_1^{(i)}, s_2^{(i)})$, $i = 1, \dots, N$. The degree of belief $Bel_{12}(B) = (Bel_1 \oplus Bel_2)(B)$ in any event B is then approximated by

$$\widehat{Bel}_{12}(B) = \frac{1}{N} \sum_{i=1}^N I\left(\Gamma_\cap(s_1^{(i)}, s_2^{(i)}) \subseteq B\right),$$

where $I(\cdot)$ is the indicator function.

III. APPLICATION TO PREDICTION

From now on, we focus on the application of belief functions to statistical prediction. The prediction of a scalar quantity, as required for one-step-ahead prediction, will first be addressed in Section III-A. Multistep prediction will then be tackled in Section III-B.

A. One-step-ahead prediction

We consider a random vector $\mathbf{X} = (X_1, \dots, X_T, X_{T+1}) \in \mathcal{X}^{T+1}$ whose joint distribution depends on some unknown parameter $\theta \in \Theta$. Let $\mathbf{X}_{1:T}$ be the vector containing the first T components of \mathbf{X} . Assume that we have observed $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$ and we want to predict X_{T+1} . The approach introduced in [1] is to predict a random variable Y with the same probability distribution as X_{T+1} given $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, of the form

$$Y = \varphi_{\mathbf{x}_{1:T}}(\theta, U), \quad (3)$$

where U is a random variable with known probability distribution, and $\varphi_{\mathbf{x}_{1:T}}$ is a mapping from $\Theta \times \mathbb{R}$ to \mathbb{R} depending on $\mathbf{x}_{1:T}$. In particular, we can define $\varphi_{\mathbf{x}_{1:T}}$ as

$$\varphi_{\mathbf{x}_{1:T}}(\theta, U) = F_{Y|\mathbf{X}_{1:T}=\mathbf{x}_{1:T}}^{-1}(U; \theta),$$

where $F_{Y|\mathbf{X}_{1:T}=\mathbf{x}_{1:T}}^{-1}(\cdot; \theta)$ is the inverse (or the generalized inverse if Y is discrete) of the cumulative distribution function of Y given $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, and U has a standard uniform distribution in $[0, 1]$.

In (3), random variable U accounts for *random* uncertainty, while lack of knowledge of θ corresponds to *epistemic* uncertainty. This latter uncertainty can be represented by a belief

function Bel_θ , which can be propagated, together with the pivotal random variable U , through (3) to obtain a belief function on Y or, equivalently, on X_{T+1} . A belief Bel_θ on θ can be defined from the likelihood function, or from a confidence region. These two methods are briefly recalled below.

a) *Likelihood-based belief function*: In [1], [2], it was proposed to consider the *likelihood-based* belief function, defined as the consonant belief function Bel_θ whose contour function is the *relative likelihood* function pl defined as

$$\pi(\theta) = pl(\theta) = \frac{L(\theta; \mathbf{x}_{1:T})}{L(\hat{\theta}; \mathbf{x}_{1:T})}, \quad (4)$$

where L denotes the likelihood function, $\hat{\theta}$ the maximum likelihood estimate (MLE) of θ , and it is assumed that $L(\hat{\theta}; \mathbf{x}_{1:T}) < \infty$. Combining Bel_θ with a Bayesian prior on θ then yields the Bayesian posterior, ensuring compatibility with Bayesian inference when a probabilistic prior is available.

b) *Confidence belief function*: Alternatively, it was proposed in [12] to model our beliefs in θ using the logical belief function Bel_θ such that $Bel_\theta(B) = I(R_{1-\alpha} \subseteq B)$, where $R_{1-\alpha}$ is a $100(1-\alpha)\%$ confidence region on θ . The only focal set of Bel_θ is then $R_{1-\alpha}$, and the corresponding contour function is the indicator function of $R_{1-\alpha}$: $\pi(\theta) = I(\theta \in R_{1-\alpha})$ for all $\theta \in \Theta$. When propagating such a *confidence belief function* in (3), we get a *calibrated* predictive belief function Bel_Y that is dominated by the true conditional distribution of Y given $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$ for $100(1-\alpha)\%$ of the samples $\mathbf{x}_{1:T}$ [12]. Under regularity conditions, an approximate $1-\alpha$ confidence region can be obtained from the relative likelihood function as

$$R_{1-\alpha} = \{\theta \in \Theta \mid pl(\theta) \geq \exp(-0.5\chi_{p;1-\alpha}^2)\}$$

where $\chi_{p;1-\alpha}^2$ is the $1-\alpha$ quantile of the chi square distribution with p degrees of freedom, and p is the dimension of θ [13].

The propagation of Bel_θ and the probability measure \mathbb{P}_U of random variable U through (3) can be formalized as a combination by Dempster's rule [2]. It can be approximated using the Monte Carlo approach described in Section II-D. The belief function Bel_θ on θ constructed by any of the two methods mentioned above is induced by a random set $([0, 1], \lambda, \beta_{[0,1]}, \Gamma)$, where Γ is defined by $\Gamma(s) = \{\theta \in \Theta \mid \pi(\theta) \geq s\}$. The corresponding predictive belief function on Y is induced by the multi-valued mapping Λ defined by

$$\Lambda : (s, u) \mapsto \Lambda(s, u) = \varphi_{\mathbf{x}_{1:T}}(\Gamma(s), u).$$

Assuming, without loss of generality, that U has a standard uniform distribution, a Monte Carlo simulation approach can be implemented by picking N pairs (s_i, u_i) , $i = 1, \dots, N$ uniformly in $[0, 1]^2$ and approximating the predictive belief function Bel_Y by the N focal sets $\Lambda(s_i, u_i)$. If these sets are real intervals, their bounds can be found by searching for the minimum and the maximum of $\varphi_{\mathbf{x}_{1:T}}(\theta, u_i)$ under the constraint $\pi(\theta) \geq s_i$ (see [1], [2]). This approach will be illustrated in Section IV-A.

B. Multiple-step-ahead prediction

Let us now consider a random vector

$$\mathbf{X} = (X_1, \dots, X_T, X_{T+1}, \dots, X_{T+h}) \in \mathcal{X}^{T+h}.$$

We assume that the first T components have been observed: $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, and the task is to predict the next h components $\mathbf{X}_{T+1:T+h}$, which have not yet been observed. As before, we can define a random vector \mathbf{Y} with the same distribution as $\mathbf{X}_{T+1:T+h}$ given $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, of the form

$$\mathbf{Y} = \varphi_{\mathbf{x}_{1:T}}(\theta, \mathbf{U}), \quad (5)$$

where \mathbf{U} is random vector with known distribution, and propagate a belief function on Θ , together with \mathbb{P}_U , through (5) (see details in [2]). However, the focal sets of the resulting predictive belief function Bel_Y are now regions of \mathbb{R}^h , for which we need to find a simple and workable representation. In this paper, we propose a solution of this problem based on the approximation of the focal sets of Bel_θ by *finite sets of points*, and their propagation in the φ -equation.

Assume that Θ is a bounded region of \mathbb{R}^p . (If Θ is not bounded, it is always possible to restrict it to a bounded region containing all parameter values θ with relative likelihood $pl(\theta)$ exceeding some threshold $\epsilon > 0$). Let $\tilde{\Theta} = \{\theta_1, \dots, \theta_M\} \subset \Theta$ be a *point cloud* [14], defined as a set of M points generated uniformly in Θ . To generate the elements of $\tilde{\Theta}$, we can use a pseudo-random number generator, but a more homogeneous covering of Θ is achieved by using a quasi-random low-discrepancy sequence such as a Halton sequence [14], [15]. As before, assume that Bel_θ is induced by a random set $([0, 1], \lambda, \beta_{[0,1]}, \Gamma)$, where Γ is defined by $\Gamma(s) = \{\theta \in \Theta \mid \pi(\theta) \geq s\}$. A focal set $\Gamma(s)$ of Bel_θ can be approximated by the set $\tilde{\Gamma}(s)$ containing the points θ in $\tilde{\Theta}$ whose plausibility $\pi(\theta)$ is greater than s :

$$\tilde{\Gamma}(s) = \{\theta \in \tilde{\Theta} \mid \pi(\theta) \geq s\}.$$

As set $\tilde{\Gamma}(s)$ is finite, it is straightforward to propagate it through the φ -equation (5) by computing $\varphi_{\mathbf{x}_{1:T}}(\theta, \mathbf{u})$ for each $\theta \in \tilde{\Gamma}(s)$. A focal set $\Lambda(s, \mathbf{u}) = \varphi_{\mathbf{x}_{1:T}}(\Gamma(s), \mathbf{u})$ of Bel_Y can then be approximated by

$$\tilde{\Lambda}(s, \mathbf{u}) = \varphi_{\mathbf{x}_{1:T}}(\tilde{\Gamma}(s), \mathbf{u}) = \{\varphi_{\mathbf{x}_{1:T}}(\theta, \mathbf{u}) \mid \theta \in \tilde{\Gamma}(s)\}.$$

Using this scheme, the predictive belief function Bel_Y can be approximated by drawing N pairs (s_i, \mathbf{u}_i) and determining the focal sets $\tilde{\Lambda}(s_i, \mathbf{u}_i)$. For any subset B of \mathcal{X}^h , the degree of belief and plausibility of B can be approximated, respectively, by

$$\widehat{Bel}(B) = \frac{1}{N} \sum_{i=1}^N I(\tilde{\Lambda}(s_i, \mathbf{u}_i) \subseteq B) \quad (6a)$$

and

$$\widehat{Pl}(B) = \frac{1}{N} \sum_{i=1}^N I(\tilde{\Lambda}(s_i, \mathbf{u}_i) \cap B \neq \emptyset). \quad (6b)$$

In the next section, we illustrate this method in the special case of a first order autoregressive model.

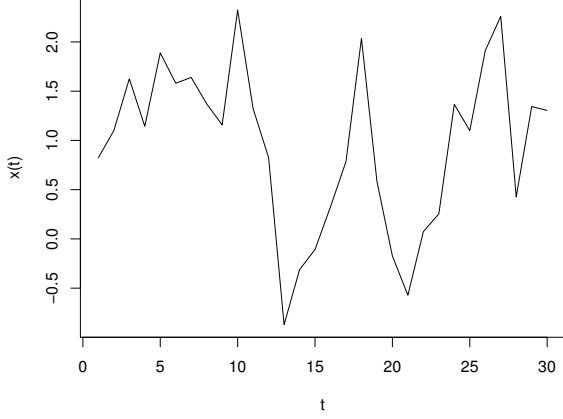


Fig. 1. Synthetic AR(1) time series.

IV. APPLICATION TO THE AR(1) MODEL

Autoregressive models (AR) are commonly used to describe time-varying phenomena in many areas such as, e.g., economics [16] or hydrology [17]. In this section, we consider, as an illustration, a zero-mean AR(1) process governed by the equation

$$X_t = \rho X_{t-1} + \epsilon_t, \quad t = 1, 2, \dots,$$

where $\rho \in (-1, 1)$ is a parameter and $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian error term with zero mean and standard deviation σ . Figure 1 shows an example of a synthetic time series of length $T = 30$ generated from this model with $\rho = 0.7$ and $\sigma = 1$.

Given that the marginal distribution of X_1 is

$$X_1 \sim \mathcal{N}\left(0, \frac{\sigma^2}{1 - \rho^2}\right)$$

and the conditional distribution of X_t given $X_{t-1} = x_{t-1}$ is $X_t | x_{t-1} \sim \mathcal{N}(\rho x_{t-1}, \sigma^2)$, the likelihood function after observing the first T terms of the sequence can easily be computed as

$$L(\theta; \mathbf{x}_{1:T}) = \phi\left(x_1; 0, \frac{\sigma}{\sqrt{1 - \rho^2}}\right) \prod_{t=2}^T \phi(x_t; \rho x_{t-1}, \sigma), \quad (7)$$

where $\theta = (\rho, \sigma)$ and $\phi(\cdot; \mu, \sigma)$ is the normal probability density function with mean μ and standard deviation σ . The likelihood-based contour function is then $pl(\theta) = L(\theta; \mathbf{x}_{1:T}) / L(\hat{\theta}; \mathbf{x}_{1:T})$, where the MLE $\hat{\theta}$ has to be computed numerically. Figure 2 shows contours of function $pl(\theta)$ for the data of Figure 1.

One-step and multiple-step ahead prediction with this model will be addressed, respectively, in Sections IV-A and IV-B.

A. One-step ahead prediction

To predict X_{T+1} , having observed $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, we introduce the following φ -equation:

$$Y = \rho x_T + \sigma \Phi^{-1}(U) = \varphi_{\mathbf{x}_{1:T}}(\theta, U),$$

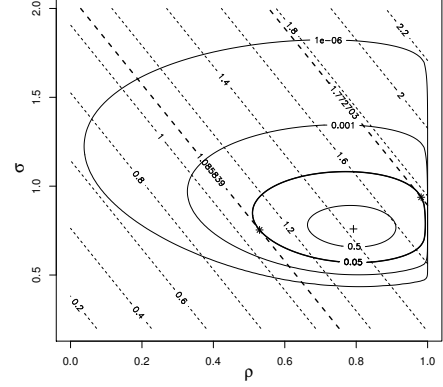


Fig. 2. Contours of the relative likelihood function $pl(\theta)$ (solid lines) and function $\varphi_{\mathbf{x}_{1:T}}(\theta, u)$ with $u = 0.7$ (broken lines), for the data of Figure 1. The MLE $\hat{\theta}$ of θ is shown as a $+$. The minimum and maximum of $\varphi_{\mathbf{x}_{1:T}}(\theta, 0.7)$ subject to $pl(\theta) \geq 0.05$ correspond to the bold broken lines, and the corresponding values of θ are shown as $*$.

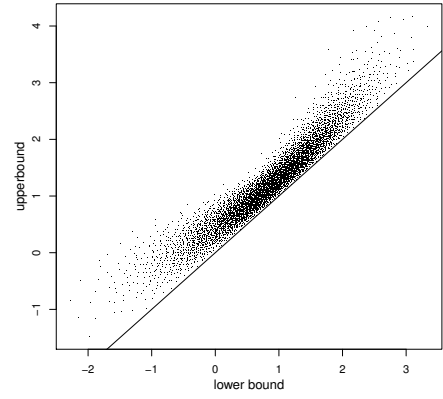


Fig. 3. Lower vs. upper bounds for $N = 5000$ focal intervals $[y_i^-, y_i^+]$ generated by picking 5000 pairs (s_i, u_i) randomly in $[0, 1]^2$.

where $Y \sim \mathcal{N}(\rho x_T, \sigma^2)$ has the same distribution as X_{T+1} given $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$ and U has a standard uniform distribution.

Figure 2 shows the contour lines of functions $pl(\theta)$ and $\varphi_{\mathbf{x}_{1:T}}(\theta, u)$ for $u = 0.7$ and the data of Figure 1. The bounds of the focal interval $[y_i^-, y_i^+] = \varphi_{\mathbf{x}_{1:T}}(\Gamma(s_i), u_i)$ for a given pair (s_i, u_i) are found by searching for the minimum and the maximum of $\varphi_{\mathbf{x}_{1:T}}(\theta, u_i)$ subject to $pl(\theta) \geq s_i$. As shown in Figure 2, for $u_i = 0.7$ and $s_i = 0.05$ we get $[y_i^-, y_i^+] = [1.09, 1.77]$. Figure 3 displays $N = 5000$ such intervals $[y_i^-, y_i^+]$ generated by picking 5000 pairs (s_i, u_i) randomly in $[0, 1]^2$. These focal intervals are a discrete representation of the predictive belief function Bel_Y . For instance, the empirical cdfs of the upper and lower bounds of intervals $[y_i^-, y_i^+]$ approximate the lower and upper cdfs of Bel_Y defined, respectively, as functions $y \mapsto Bel((-\infty, y])$ and $y \mapsto Pl((-\infty, y])$ (see Figure 4).

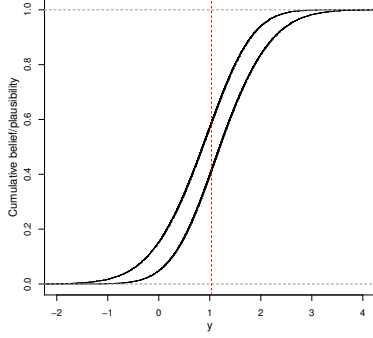


Fig. 4. Approximations of the lower and upper cdfs of Bel_Y computed using the 5000 focal intervals $[y_i^-, y_i^+]$ displayed in Figure 3. The vertical broken line corresponds to the plug-in prediction $\hat{x}_{T+1} = \hat{\rho}x_T$.

B. Multistep prediction

We now consider multistep prediction. For instance, for $h = 3$ steps ahead, we have

$$\begin{aligned} x_{T+1} &= \rho x_T + \epsilon_{T+1} \\ x_{T+2} &= \rho x_{T+1} + \epsilon_{T+2} \\ x_{T+3} &= \rho x_{T+2} + \epsilon_{T+3}, \end{aligned}$$

where ϵ_{T+1} , ϵ_{T+2} and ϵ_{T+3} are independent random variables with normal distribution $\mathcal{N}(0, \sigma^2)$. Given $X_T = x_T$, the random vector $\mathbf{X}_{T+1:T+3} = (X_{T+1}, X_{T+2}, X_{T+3})'$ has the same distribution as

$$\mathbf{Y} = \varphi_{\mathbf{x}_{1:T}}(\theta, \mathbf{U}) = (Y_1, Y_2, Y_3)'$$

with

$$\begin{aligned} Y_1 &= \rho x_T + \sigma \Phi^{-1}(U_1) \\ Y_2 &= \rho Y_1 + \sigma \Phi^{-1}(U_2) \\ Y_3 &= \rho Y_2 + \sigma \Phi^{-1}(U_3), \end{aligned}$$

where $\mathbf{U} = (U_1, U_2, U_3)$, and U_1 , U_2 and U_3 are independent random variables with a standard uniform distribution $\mathcal{U}([0, 1])$.

As explained in Section III-B, the focal sets $\Lambda(s, \mathbf{u}) = \varphi_{\mathbf{x}_{1:T}}(\Gamma(s), \mathbf{u})$ of Bel_Y are now regions of \mathbb{R}^3 , which will be represented by finite sets of points. For that purpose, we start by representing the parameter space Θ by a set $\tilde{\Theta}$ of $M = 3 \times 10^4$ points, as shown in Figure 5a. A focal set $\Gamma(s) = \{\theta \in \Theta \mid pl(\theta) \geq s\}$ of Bel_θ is then represented by a finite set of points $\tilde{\Gamma}(s) = \{\theta \in \tilde{\Theta} \mid \pi(\theta) \geq s\}$ (see Figure 5b). For any realization \mathbf{u} of \mathbf{U} , we can then approximate the focal set $\Lambda(s, \mathbf{u})$ by computing the image $\varphi_{\mathbf{x}_{1:T}}(\theta, \mathbf{u})$ of each $\theta \in \tilde{\Gamma}(s)$. We then obtain the set $\tilde{\Lambda}(s, \mathbf{u}) = \varphi_{\mathbf{x}_{1:T}}(\tilde{\Gamma}(s), \mathbf{u})$. The two-dimensional projections of $\tilde{\Lambda}(s, \mathbf{u})$ on the planes spanned by (Y_1, Y_2) and (Y_2, Y_3) for three different randomly picked values of \mathbf{u} are shown, respectively, in Figures 5c and 5d. The convex hulls of 100 focal sets $\tilde{\Lambda}(s_i, \mathbf{u}_i)$ are shown in Figure 6.

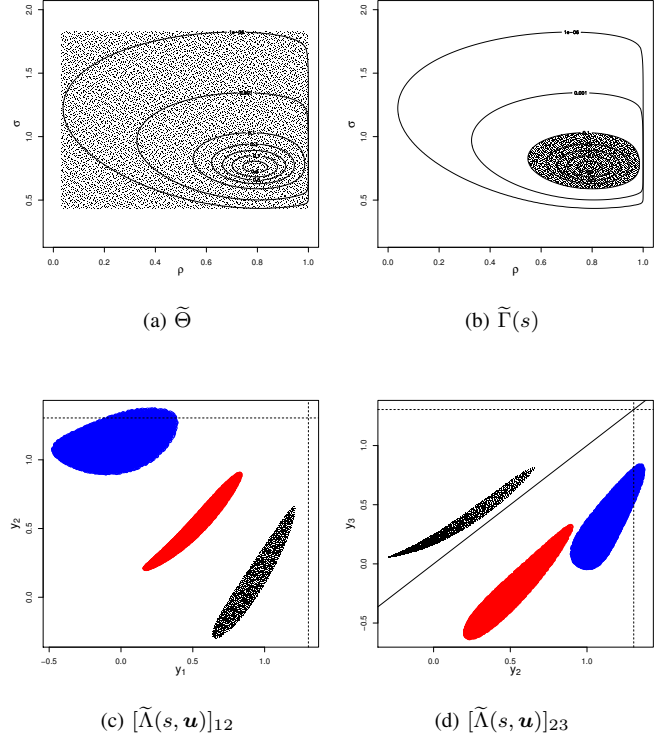


Fig. 5. Illustration of the approximation of focal sets using point clouds. (a): Approximation of parameter space Θ by a finite set $\tilde{\Theta}$ of M points; (b): Approximation of a focal set $\Gamma(s)$ of Bel_θ for $s = 0.1$; (c) and (d): 2-D projections of focal sets $\tilde{\Lambda}(s, \mathbf{u})$ for $s = 0.1$ and three different values of \mathbf{u} .

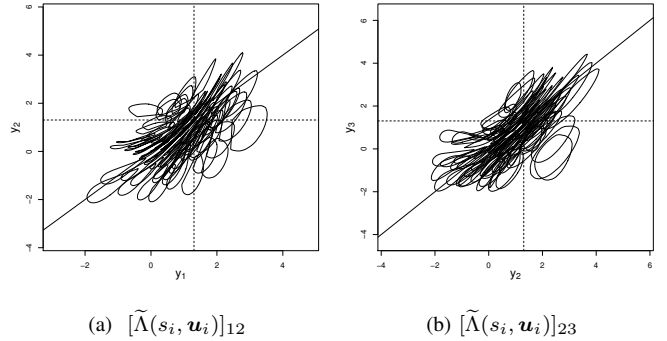


Fig. 6. Convex hulls of the two-dimensional projections of 100 focal sets $\tilde{\Lambda}(s_i, \mathbf{u}_i)$ on the planes spanned by (Y_1, Y_2) (left) and (Y_2, Y_3) (right).

After N focal sets $\tilde{\Lambda}(s_i, \mathbf{u}_i)$, $i = 1, \dots, N$ have been generated, they can be used to approximate the degrees of belief and plausibility of events of interest using (6). For instance, let us consider the following three events:

$$\begin{aligned} E_1 &= (X_{T+1} > X_{T+2} > X_{T+3}) \\ E_2 &= (X_{T+1} < 0) \text{ and } (X_{T+2} < 0) \text{ and } (X_{T+3} < 0) \\ E_3 &= (X_{T+1} > 0) \text{ and } (X_{T+2} > 0) \text{ and } (X_{T+3} > 0). \end{aligned}$$

Table I shows the degrees of belief and plausibility of these

TABLE I

BELIEF, PLAUSIBILITY AND PROBABILITY OF THREE EVENTS. THE BELIEF-PLAUSIBILITY INTERVALS WERE COMPUTED BASED ON THE LIKELIHOOD-BASED AND CONFIDENCE BELIEF FUNCTIONS ON Θ .

Event	Likelihood		Confidence		P
	Bel	Pl	Bel	Pl	
E_1	0.26	0.38	0.19	0.42	0.31
E_2	0.024	0.073	0.011	0.11	0.086
E_3	0.51	0.75	0.38	0.83	0.50

TABLE II

ESTIMATED COVERAGE PROBABILITIES OF BELIEF-PLAUSIBILITY INTERVALS COMPUTED FROM THE LIKELIHOOD-BASED AND CONFIDENCE BELIEF FUNCTIONS ON θ .

	E_1	E_2	E_3
Likelihood	0.987	0.924	0.782
Confidence	0.998	0.996	0.984

three events, as well as their true conditional probability given $X_T = x_T$ estimated by Monte Carlo simulation. The belief-plausibility intervals were computed using both the likelihood-based function and a 95% confidence belief function on Θ (see Section III-A). We recall that, using the latter, we have, for at least 95% of the samples, $Bel_Y(E) \leq P(E | X_T = x_T) \leq Pl_Y(E)$ for any event E [12]. In contrast, these inequalities are not guaranteed to hold when the likelihood-based belief function is used. As shown in Table I, the belief-plausibility intervals computed from the confidence belief function do contain the true probabilities of the three events for this particular dataset, as expected. The likelihood-based intervals are narrower, but they do not have frequency-calibration properties.

To estimate the coverage probability of the belief-plausibility intervals (i.e., the probability of having $Bel_Y(E) \leq P(E | X_T = x_T) \leq Pl_Y(E)$, computed over a large number of realizations of the time series), we repeated the calculation with 1000 times series generated from the same distribution. The value of x_T was fixed at 1.304641, as in the previous calculations. The estimated coverage probabilities are reported in Table II. As expected, the belief-plausibility intervals based on the confidence belief function all have a coverage probability greater than 95%, whereas the likelihood-based intervals have coverage probabilities below the nominal 95% level for events E_2 and E_3 .

V. CONCLUSIONS

Although belief functions were initially introduced for statistical inference [3], they have not been widely used in statistical applications due to the difficulty to represent and manipulate continuous belief functions in \mathbb{R}^p . In this paper, we have proposed a method to represent and propagate continuous belief functions in multidimensional spaces based on the approximation of focal sets by point clouds. We have applied this method to the prediction of random vectors and, more specifically, to multistep time series prediction, extending the

applicability of the prediction method introduced in [1]. The same method can be applied to a wide range of problems, such as, e.g., uncertainty propagation in numerical equations describing systems with uncertain parameters and/or inputs.

A key advantage of the point-cloud representation is the possibility to approximate any region of \mathbb{R}^p with any accuracy, provided enough points are used. Furthermore, basic set-theoretic operations such as projection, intersection, testing inclusion, etc., can be performed very efficiently. In contrast to the polygon representation recently introduced in [18], the point-cloud representation can be applied with any number p of dimensions, but the curse of dimensionality obviously limits its use to spaces of moderate values of p , as the number of points needed to approximate a region with some accuracy grows exponentially with p . More simulations are needed to study the influence of p and the number M of points on approximation accuracy.

REFERENCES

- [1] O. Kanjanatarakul, S. Sriboonchitta, and T. Denœux, "Forecasting using belief functions: an application to marketing econometrics," *International Journal of Approximate Reasoning*, vol. 55, no. 5, pp. 1113–1128, 2014.
- [2] —, "Statistical estimation and prediction using belief functions: principles and application to some econometric models," *International Journal of Approximate Reasoning*, vol. 72, pp. 71–94, 2016.
- [3] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.
- [4] G. Shafer, *A mathematical theory of evidence*. Princeton, N.J.: Princeton University Press, 1976.
- [5] T. Denœux, "Likelihood-based belief function: justification and some extensions to low-quality data," *International Journal of Approximate Reasoning*, vol. 55, no. 7, pp. 1535–1547, 2014.
- [6] H. T. Nguyen, "On random sets and belief functions," *Journal of Mathematical Analysis and Applications*, vol. 65, pp. 531–542, 1978.
- [7] G. Shafer, "Allocations of probability," *Annals of Probability*, vol. 7, no. 5, pp. 827–839, 1979.
- [8] G. de Cooman and D. Aeyels, "A random set description of a possibility measure and its natural extension," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 2, pp. 124–130, 2000.
- [9] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, pp. 3–28, 1978.
- [10] G. Shafer, "Dempster's rule of combination," *International Journal of Approximate Reasoning*, vol. 79, pp. 26–40, 2016.
- [11] N. Wilson, "A Monte Carlo algorithm for Dempster-Shafer belief," in *Proc. of the 7th Conference on Uncertainty in AI*, B. D'Ambrosio, P. Smets, and P. Bonissone, Eds. Morgan Kaufmann, 1991.
- [12] T. Denœux and S. Li, "Frequency-calibrated belief functions: Review and new insights," *International Journal of Approximate Reasoning*, vol. 92, pp. 232–254, 2018.
- [13] Y. Pawitan, *In all likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Clarendon Press, 2001.
- [14] L. Sui, P. Feissel, and T. Denœux, "Identification of elastic properties in the belief function framework," *International Journal of Approximate Reasoning*, vol. 101, pp. 69–87, 2018.
- [15] J. H. Halton, "Algorithm 247: Radical-inverse quasi-random point sequence," *Communications of the ACM*, vol. 7, no. 12, pp. 701–702, 1964.
- [16] W. H. Greene, *Econometric analysis*, 7th ed. Upper Saddle River, NJ, USA: Prentice Hall, 2012.
- [17] D. Machiwal and M. K. Jha, *Hydrologic Time Series Analysis: Theory and Practice*. Springer, 2012.
- [18] N. Pellicanò, S. L. Hégarat-Masclé, and E. Aldea, "2CoBel: a scalable belief function representation for 2D discernment frames," *International Journal of Approximate Reasoning*, vol. 103, pp. 320–342, 2018.