Luceños' Discretization Method and its Application in Decision Making under Ambiguity

Michael Obermeier
Department of Statistics
University of Munich
Germany
obermeierm@web.de

Thomas Augustin
Department of Statistics
University of Munich
Germany
Thomas@stat.uni-muenchen.de

Abstract

When extending classical statistical models to imprecise probabilities, one fundamental difficulty, which may have hindered some powerful practical applications, is the following gap: While classical statistical models are typically based on absolutely continuous probability distributions, most computational methods developed for handling imprecise probability models rely on finite sample spaces. A natural way to close this gap is discretization of the underlying continuous probability distribution. This, however, is far from straightforward, because naïve discretization by mere rounding may cause a substantial bias; even moments of very low order would be distorted. The present paper discusses the application of Luceños' ([10]) so-to-say adaptive discretization method in imprecise probability models. We firstly recall two theorems, showing, for any fixed natural number r, how to construct a discrete random variable such that its first, second, ... r-th moment coincides with the corresponding moment of the underlying continuous distribution. (In addition, also coincidence of the distribution functions in a fixed number of points can be enforced.) Then we illustrate the power of the method by utilizing it in decision problems under ambiguity.

Keyword: Decision making under ambiguity, discretization, Gaussian quadrature, imprecise probabilities, interval probability, linear programming, Luceño, numerical integration.

1 Introduction

Classical statistical models typically are based on parametric, absolutely continuous probability distributions on the real line. Handling extensions of these models in the imprecise probability framework, quite often becomes very demanding from the computational point of view, and then approximative techniques are the best one can hope for, the more as also in classical statistics many integrals of less smooth functions can be only obtained numerically. A natural idea in this context is discretization, in order to make available powerful algorithms (for instance for handling graphical models ([4]) or decision making ([9], [17]) that explicitly rely on finite spaces to obtain approximate solutions in this generalized setting. However, such discretizations need some care; for more than hundred years, since the work of Sheppard ([13]) at the end of the nineteenth century, statisticians have been well aware that analysis based on rounded data may be severely biased, and so discretization by mere rounding or other ad-hoc techniques is a bad advice. Sheppard also developed a simple correction formula, the applicability of which, however, is restricted to the r-th moment of a normal distribution and further regularity conditions. Applying this "correction" to other distributions or more complex functions of random variables can even increase the bias ([18], [1]).

In this paper we study a more sophisticated discretization procedure, which discretizes the underlying distribution in an adaptive way such that, for any r, the r-th first moments of the new, discretized distribution agree with the ones from the original distribution. The method is based on Gaussian quadrature; its power in statistical applications is advocated by Luceño ([10]), to which we refer when describing the essentials of the technique. We claim that these results are even more important in the area of imprecise probabilities, for the reasons outlined above. To corroborate this thesis, we illustrate the discretization technique by applying it to two types of decision problems under ambiguity.

In more detail, the paper is organized as follows: In Section 2 we prepare the ground by recalling Luceños ([10]) two main theorems and briefly discussing issues of concrete calculation of the discretization. Our application to decision theory is divided in four sections. In Section 3 we recall basic notions, which then are used to explain in Section 4 two different general types of discretizations. In Section 5 it is shown how to use

one discretization technique in parametric situations with non-elementary utility functions: prior information is assumed to be described by a set of parametric distributions with varying parameters. We calculate E-admissible actions as well as the Γ-maximin solutions. While these results may be understood as a direct extension of numerical integration procedures from classical probability to the imprecise probability framework, the importance of the second type of discretization, the construction of an approximately equivalent decision problem becomes unambiguously evident in Section 6, where envelopes of parametric sets are considered. After choosing a reference distribution which determines the concrete discretization, we transform the whole decision problem on an infinite state of natures to an approximately equivalent decision problem based on a finite set of states of nature, and then give E-admissible and Γ-maximin solutions by adapting the algorithms from [9] and [17].

2 Luceños Discretization Method

2.1 Two Fundamental Theorems

One can understand Luceños article *discrete approximations to continuous univariate distributions* ([10]) as a kind of "translation" of the Gaussian quadrature method for integration into probability theory.

The central idea of the Gaussian quadrature method is to replace an integral over a function $h(x)$ and its weight function $w(x)$ by a sum, i.e. to take

$$\int_a^b h(x)w(x)d(x) \approx \sum_{j=1}^{N} h(x_j)w_j, \tag{1}$$

where w_j and x_j are chosen in a sophisticated way such that one can approximate the value of the integral numerically with rather high accuracy. To find the nodes x_j and the weights w_j, $j = 1, \ldots, N$, the roots of recursively defined orthogonal polynomials of degree N are used, which depend on $w(x)$. Unlike some other numerical integration methods, e.g. the Newton-Cotes formulas, the abscissae and weights here are found dynamically, which means that they are adapted to the shape of the original function and therefore the approximation of the integral is very accurate (for Gaussian quadrature and other numerical integration methods see for example [15]).

In the following statistical application of this method the weight function $w(x)$ is the probability density function (PDF) of a univariate continuous random variable X, whereas w_i, $i = 1, \ldots, N$ constitute the probability mass function (PMF) of the corresponding discrete random variable Y. The detailed proceeding and some fundamental properties of the approximations are described in two theorems, which we recall here from [10].

Proposition 1 (Luceño) *Consider two univariate random variables X and Y on a domain $[a, b]$ with $-\infty \leq a \leq x \leq b \leq \infty$.[1] Let X be (absolutely) continuous with probability density function (PDF) $w(x)$, having finite moments of any order, and let Y be discrete on N atoms $x_1, \ldots, x_N$ with probability mass function (PMF) $w_1, \ldots, w_N$. Then*

$$\mathbb{E}(X^r) = \mathbb{E}(Y^r), \forall r \in \{0, \ldots, 2N-1\} \tag{2}$$

if and only if the nodes $x_1, \ldots, x_N$ *and the* weights $w_1, \ldots, w_N$ *satisfy the following two conditions:*

i) $x_1, \ldots, x_N$ are the roots of the polynomial $Q_N(x)$ of degree N defined by the three-term recursion

$$Q_{i+1}(x) = (x - \delta_{i+1})Q_i(x) - \gamma_{i+1}^2 Q_{i-1}(x), \tag{3}$$

$$i \geq 0, \text{ where } Q_{-1}(x) \equiv 0,\ Q_0(x) \equiv 1 \text{ and}$$

$$\delta_{i+1} = \mathbb{E}\{XQ_i^2(X)\}/\mathbb{E}\{Q_i^2(X)\},\ i \geq 0, \tag{4}$$

$$\gamma_{i+1}^2 = \begin{cases} 0, & i = 0 \\ \mathbb{E}\{Q_i^2(X)\}/\mathbb{E}\{Q_{i-1}^2(X)\},\ i \geq 1. \end{cases} \tag{5}$$

ii) the probabilities $w_1, \ldots, w_N$ are the solution of the linear system

$$\sum_{j=1}^{N} Q_k(x_j)w_j = \begin{cases} 1, & k = 0 \\ 0, & k = 1, \ldots, N-1. \end{cases} \tag{6}$$

In addition, the cumulative distribution functions (CDFs) of X and Y can be forced to agree at least in a given set of points:

Proposition 2 (Luceño) *Consider the situation of Proposition 1. Let $c_0 = a < c_1 < \ldots < c_{M-1} < c_M = b$ such that $I_i = \int_{c_{i-1}}^{c_i} w(x)dx > 0$ for all $i = 1, \ldots, M$, and consider the discrete random variables Y_i, $i = 1, \ldots, M$ with atoms $x_{i1}, \ldots, x_{iN}$ and weights $w_{i1}, \ldots, w_{iN}$ arising from applying Proposition 1 to the random variables $X_i := X \cdot 1_{[c_{i-1}, c_i]}$.*

Then, for the random variable Z with atoms $x_{11}, \ldots, x_{1N}, x_{21}, \ldots, x_{2N}, \ldots, x_{M1}, \ldots, x_{MN}$ and weights $I_1 \cdot w_{11}, \ldots, I_1 \cdot w_{1N}, I_2 \cdot w_{21}, \ldots, I_2 \cdot w_{2N}, \ldots, I_M \cdot w_{M1}, \ldots, I_M \cdot w_{MN}$,

$$\mathbb{E}(X^r) = \mathbb{E}(Z^r), \quad \forall r \in \{0, \ldots, 2N-1\}, \tag{7}$$

and the CDFs of X and Y coincide at least at the abscissae $c_0, c_1, \ldots, c_{M-1}, c_M$.

[1] To include the statistical standard distributions, without the need to distinguish between the domain $\mathbb{R}$ and some bounded domain, we allow for $a = \infty$ and $b = \infty$, but implicitly assume $f(-\infty) = f(\infty) = 0$

2.2 Easier Calculations in Standard Cases

In the case where the weight function belongs to a standard family, there are well known polynomials, which can be used instead of the three-term recursion just described to find the abscissae and corresponding weights. For example, for the normal distribution with mean μ and variance σ^2 one can use the Gauss-Hermite polynomials with the weight function $w(x) = e^{-x^2}$ on the interval $-\infty < x < \infty$:[2]

$$H_{j+1}(x) = 2xH_j(x) - 2jH_{j-1}(x) \qquad \text{with} \quad H_1(x) = 1$$

Given the roots $x_i^{(GH)}$ and weights $w_i^{(GH)}$ of the polynomial with degree N, one can obtain the random Variable Y and its PMF through (see [10], p.347):

$$Y_j = \mu + \sigma x_j^{(GH)}\sqrt{2}, \quad w_j = \frac{1}{\sqrt{\pi}} w_j^{(GH)}, j = 1, \ldots, N.$$

However there are only very few, classical families where one can easily use well known polynomials to find the new variable.

2.3 Nonstandard Case

For the partial intervals used in Proposition 2 the weights do not generally belong to any of these classical families. As a consequence, there are no well known polynomials which can be used to find the new variable, and the three term recursion described above has to be used, leading to another numerical problem: how to solve the inner products in Part i) of Proposition 1 to determine the γ_is and δ_is, if $w(x)$ is no classical weight?

Based on the knowledge of the so called "modified moments" $\nu_j = \int_a^b \pi_j(x)w(x)dx$ of orthogonal polynomials π_j, Sack and Donavan ([12]) offer a numerically stable algorithm to find the coefficients γ_j^2 and δ_j of the recursion. Wheeler ([22]) improved this method to an $O(N^2)$ algorithm. Other solutions are presented by Gautschi ([6]). One simple and heuristic way is to approximate the inner products with an adequate quadrature rule. In the further calculations presented here this simple method is used, because the focus of this paper is mostly on the construction of the new discrete random variable, not directly on the exact calculation of an integral.

In a last step the weights w_i and nodes x_i of the new variable have to be found. They result from the eigenvalues and eigenvectors of a tridiagonal matrix consisting of the γ_i^2 and δ_i from (5) and (4) (for details see [15], p.179f.).

2.4 Accuracy of the Approximation

If the focus of the approximation is on the shape of a continuous CDF, one will use the method described in Proposition 2. The accuracy of this discrete approximation depends on the way the partition of $[a, b]$ is chosen. One first possibility is to split the support in inner intervals $[c_{i-1}, c_i]$, $i = 2, \ldots, M-1$ of the same size and two possibly larger outer intervals if the domain is infinite. The obvious problem is that then one has the same number of interpolation points in areas where the PDF is high (which means that the CDF has a big increase) as in areas with low PDF. A more satisfying method is to use the PDF (if it is numerically manageable) to find a more appropriate partition. One can split the support in M quantiles and use them as the c_j's. In this way the intervals are adjusted to the shape of the original distribution: they are small where the PDF is high and wide where the PDF is low, and so finally in the important areas of the support there are more nodes than in the less important ones.[3]

3 Decision Making under Ambiguity, Basic Notions

To illustrate and exemplify the power of Luceños method in the area of imprecise probabilities, we apply it to some general decision problems under ambiguity. To prepare the ground we briefly recall the basic setting of decision theory, where one has to choose an optimal *action* from a non-empty, finite set $\mathbb{A} = \{a_1, ..., a_n\}$ of possible actions. The consequences of every action depend on the true, but unknown *state of nature* ϑ being an element of a space Θ. The corresponding outcome is evaluated by the *utility function* $u : (\mathbb{A} \times \Theta) \to \mathbb{R}$ and by the associated random variable $\mathbf{u}(a)$ on Θ. Often it makes sense to study *randomized actions* in addition, which can be understood as a classical probability measure $\lambda = (\lambda_1, ..., \lambda_n)$ on $(\mathbb{A}, \mathcal{P}o(\mathbb{A}))$, where λ_i is interpreted as the probability with which action a_i is taken. Then $u(\cdot)$ and $\mathbf{u}(\cdot)$ are extended to randomized actions by defining $u(\lambda, \vartheta) := \sum_{s=1}^n u(a_s, \vartheta)\lambda_s$. (Next to simplifying calculations, under some criteria the optimal randomized action may be superior to the optimal unrandomized one.)

[2] For some other distributions there are also well known polynomials, which can be used directly to find the discretization: For the gamma distribution the Gauss-Laguerre polynomials can be used, for the beta distribution the Gauss-Jacobi polynomials and for the uniform distribution the Gauss-Legendre polynomials (for details see [10]).

[3] Several tests in [11], chapter 3 for this approximation to the standard normal distribution show empirically that, for a sufficiently large M, samples of the new discrete variable cannot be distinguished any more from the original variable.

This model contains the essentials of every (formalized) decision situation under uncertainty and is applied in a huge variety of disciplines. If the states of nature are produced by a perfect random mechanism, and if the corresponding probability measure $\pi(\cdot)$ on $(\Theta, \mathcal{P}o(\Theta))$ is completely known, the Bernoulli principle is nearly unanimously favored. One chooses then the unrandomized action a^* or the randomized action λ^* maximizing the expected utility

$$\mathbb{E}_\pi \mathbf{u}(a) := \int u(a, \vartheta) d\pi(\vartheta) \tag{8}$$

and $\mathbb{E}_\pi \mathbf{u}(\lambda) := \int u(\lambda, \vartheta) d\pi(\vartheta)$ among all a and all λ, respectively. a^* and λ^*, respectively, is called *Bayes action with respect to* π. In many applications, however, it is not possible to describe the prior knowledge on the stochastic behavior of the states of nature by a classical probability measure, and a more general description of ambiguity is needed, as provided by imprecise probabilities and related approaches (see, in particular, [19] and [21]).

From the technical point of view, the usual concepts of imprecise probability lead to convex sets $\mathcal{M}$ of classical probabilities. Every distribution π from $\mathcal{M}$ produces a classical expected utility $\mathbb{E}_\pi \mathbf{u}(\lambda)$. Assuming $\mathcal{M} \neq \varnothing$, all possible expected utilities $\mathbb{E}_\pi \mathbf{u}(\lambda)$ range within the interval

$$\left[\underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda) \, , \, \overline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda) \right] , \tag{9}$$

and this interval-valued quantity is called *generalized expected utility*. Based on this notion of generalized expected utility several optimality criteria are common. An overview is given in [16] where also further references are provided. Here two of them are considered: the Γ-maximin criterion and the criterion of E-admissibility.

The Γ-*maximin* criterion considers a worst case scenario, which means $\underline{\mathbb{E}}_{\mathcal{M}} u(\lambda)$ is evaluated only. Then an action λ^* is optimal iff for all λ

$$\underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda^*) \geq \underline{\mathbb{E}}_{\mathcal{M}} \mathbf{u}(\lambda). \quad {}^{4} \tag{10}$$

The concept of *E-admissibility* on the other hand typically does not offer a unique action as the one to choose, but a set of optimal actions: An action[5] a^* is said to be E-admissible in $\mathcal{M}$ with respect to a set of prior probabilities $\mathcal{M}$, iff there exists a classical prior $\pi(\cdot) \in \mathcal{M}$ such that a^* is Bayes action with respect to $\pi(\cdot)$ for all actions a under consideration.[6]

4 Two Types of Discretization

In the case of continuous distributions of the states of nature practical handling these criteria may encounter severe difficulties. Except in special cases, where the distributions are stochastically ordered and or the expected utility is easily expressed by a underlying parameter, it is hard or even impossible to determine optimal actions by evaluating the integrals (8). Since for finite set of states of nature powerful algorithms exist, decision making provides an area where Luceños discretization techniques is quite welcome.

To implement the discretization, we apply Proposition 1 and 2 by assuming there is a random variable X. In the background, that takes values in Θ producing the states of natures. Applying the general techniques to this variable X setting in (1) as well as to $w(\cdot) = \pi(\cdot)$ and $h(\cdot) = u(a, \cdot)$ and $h(\cdot) = u(\lambda, \cdot)$, respectively, note that not only the weights but also the nodes depend on the underlying probability distribution. Therefore, for a given set $\mathcal{M}$ of continuous distributions on $\Theta \subset \mathbb{R}$ (equipped with the corresponding Borel $\sigma-$field $\mathcal{B}$), two different types of discretization have to be distinguished, depending whether a separate or a common discretization scheme is used:

Type-I discretization: The first possibility is to apply Proposition 1 or 2 to every element $\pi(\cdot) \in \mathcal{M}$ separately. This means that to every $\pi(\cdot) \in \mathcal{M}$ a corresponding discrete distribution $\nu_\pi(\cdot)$ is constructed with atoms $\vartheta_{1,\pi}, \ldots, \vartheta_{N,\pi}$, and $\mathbb{E}_\pi \mathbf{u}(\lambda)$ is replaced by its approximative equivalent

$$\mathbb{E}_{\nu_\pi} \mathbf{u}(\lambda) = \sum_{j=1}^{N} u(\lambda, \vartheta_{j,\pi}) \nu_\pi \left(\vartheta_{j,\pi} \right). \tag{11}$$

Type-II discretization: Here Θ itself is discretized. For this, a certain *reference distribution* $\pi_0(\cdot) \in \mathcal{M}$ is selected, to which then Proposition 1 or 2 is applied.[7] The resulting nodes[8] $x_1, \ldots, x_N$ are used to define a new discrete space $\Theta_d = \{\theta_1, \ldots, \theta_N\}$ with $\vartheta_1 = \big[a, (x_1 + x_2)/2\big]$, $\vartheta_2 = \big((x_1 + x_2)/2, (x_2 + x_3)/2\big]$, $\ldots$, $\vartheta_N = \big((x_{N-1} + x_N)/2, b\big]$, which then is used to replace Θ. The utility function is then extended to Θ_d

[4] This concept is a very conservative decision rule, similarly to the maximin rule in the classical decision theory: in the case of complete ambiguity both criteria coincide.

[5] Under the criterion of E-admissibility usually consideration is confined to the unrandomized actions. If needed, the algorithms used later on can be extended to randomized actions (cf. [17, p. 357]).

[6] E-admissibility can be considered in a broader sense as a generalization of the criterion of admissibility in classical decision theory.

[7] This method is described later on in Section 6. A glance at Figure 4 may therefore be helpful.

[8] For the sake of readability, the dependence on $\pi_0(\cdot)$ is suppressed in the notation here throughout the following definitions.

by assigning the values at the nodes, i.e. by defining

$$u(a, \vartheta_j) := u(a, x_j), \quad \forall j \in \{1, \ldots, N\}, \ \forall a \in \mathbb{A}\,.$$

The nodes of the reference distribution are used to discretize all elements of $\mathcal{M}$. More precisely, the set $\mathcal{M}$ of continuous probability distributions on $(\Theta, \mathcal{B})$ is replaced by the set $\mathcal{P}$ of discrete probability distributions, being the set of all classical probabilities in accordance with $P(\cdot) = [L(\cdot), U(\cdot)]$ on $(\Theta_d, \mathcal{P}o(\Theta_d))$[9] defined via:[10]

$$L\Big(\bigcup_{j\in J}\{\vartheta_j\}\Big) = \inf_{\pi\in\mathcal{M}} \pi\Big(\bigcup_{j\in J}\vartheta_j\Big) \tag{12}$$

and

$$U\Big(\bigcup_{j\in J}\{\vartheta_j\}\Big) = \sup_{\pi\in\mathcal{M}} \pi\Big(\bigcup_{j\in J}\vartheta_j\Big) \tag{13}$$

$$\forall J \subset \{1, \ldots, N+1\}\,. \tag{14}$$

Since N may be quite large, in many applications the computational effort may be substantially reduced by a further approximation in which not all elements of $\mathcal{P}o(\Theta_d)$ are used in the assignment of $P(\cdot)$. Then the power set of $\{1, \ldots, N+1\}$ in (14) is replaced by some subset $\mathcal{J}_{\max}$, and natural extension is applied to obtain the remaining interval limits $L(\cdot)$ and $U(\cdot)$. A natural choice for $\mathcal{J}_{\max}$, that is also applied in Section 6, is to consider only connected intervals in the assignment procedure (cf. also Figure 4).

Both types of discretization have different types of applications. Type-I discretization necessarily requires that the density functions of all elements of $\mathcal{M}$ have to be known, in order to able to apply Luceño's theorems to each of them. In particular, the set $\mathcal{M}$ must be dominated by the Lebesgue measure (in the measure-theoretic sense) to guarantee the existence of appropriate densities. These conditions do not apply for the second option, which therefore is more general. There it is sufficient that the reference distribution has a known density, the set $\mathcal{M}$ itself may even be undominated, which is usually for instance the case when considering neighborhood models from robust statistics (e.g., [7], [3], [20], [14]).

We will discuss both methods, from the general point of view as well as with the help of examples. For ease of illustration we will use in both examples a set of normal distributions. In Case i) it is used immediately as the credal prior information $\mathcal{M}$, in Case ii) it serves as a building block to define an appropriate interval-valued assignment.

[9] $\mathcal{P}o(\Theta_d)$ denotes the power set of Θ_d.

[10] By conjugacy ($U(\cdot) = 1 - L(\cdot^C)$) either (12) or (13) would be sufficient to describe $P(\cdot)$.

5 Applications of Type-I Discretization

We start with Type-I discretization, where $\mathcal{M}$ is a set of absolutely continuous probability distributions, to each of which the discretization procedure is applied. We assume that $\mathcal{M}$ can be described by a set $(f(\cdot)_\psi))_{\psi\in\Psi}$ of densities with the parameter space Ψ being a compact subset of $\mathbb{R}^k$ for some finite k. In this situation, for every $\pi(\cdot) \in \mathcal{M}$, the expected utility of an action can be (approximately) calculated by Equation (11), relying on the new discrete distribution ν_π. The optimal action (the action with the largest expected utility) can be found with linear optimization. It can be seen as the value of a function depending on the unknown parameter ψ. When n, the number of competing actions is small, as in the following example, E-admissible actions as well as the Γ-maximin action can be extracted graphically by plotting these functions. Section 5.2 then sketches general computational tools for complexer situations with larger n.

5.1 Numerical example

In the following the procedure should be firstly explained with the help of a numerical example. Consider the actions a_1, a_2 and a_3 with their utility functions

$$\begin{aligned} u(a_1, \vartheta) &= \exp(-\exp(\vartheta)) \\ u(a_2, \vartheta) &= \exp(-\exp(\vartheta^2)) \\ u(a_3, \vartheta) &= 0.1\,. \end{aligned}$$

The associated state of nature follow a normal distribution with $\mu = 1$ and σ varying between 0.5 and 1.5. This is an example where it is impossible to solve the corresponding integrals of the expected utility analytically, and we have to rely on Luceños method.[11]

Relying on the criterion of E-admissibility, the results can be found in Figure 1. As we are interested in the expected utility of an action in dependence on σ, which means in the value of an integral, it is reasonable to use here the simple Gaussian quadrature rule from Proposition 1 for the discretization.[12] On the left hand side one can identify the optimal action in dependence on σ, while on the right hand side the corresponding expected utility of the optimal action is shown. Two of the three actions are optimal for special values of σ; the set of E-admissible actions is $\{a_1, a_2\}$.

[11] Such integrals for instance occur when handling frailty or measurement error in survival models. Note further that with the first utility function even such common techniques like Taylor series expansion fail to calculate the integral.

[12] For Figure 1 a discretization with $N = 10$ points is used for the normal distribution with $\mu = 1$ and any fixed σ.

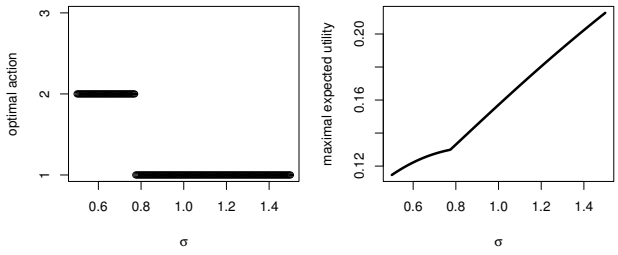

Figure 1: Action with maximal expected utility depending on σ (left), maximal possible utility depending on σ (right)

If additionally the mean μ is uncertain, the two-dimensional Figure 1 becomes three-dimensional (cf. Figure 2). The parameter μ now also varies in

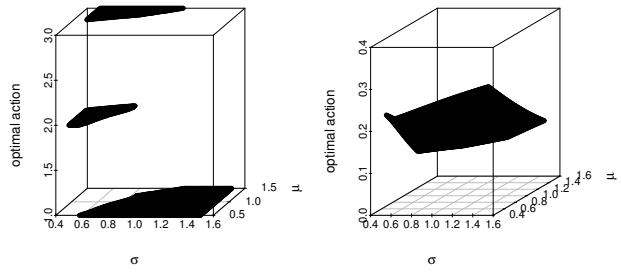

Figure 2: Action with maximal expected utility depending on σ and μ (left), maximal possible utility depending on σ and μ (right)

the interval $[0.5, 1.5]$. Like in the picture before, one can identify on the left hand side for a special μ-σ combination the optimal action. On the right there are again the corresponding expected utilities. For some μ-σ combinations now also action a_3 is optimal, which means that the set of E-admissible actions consists of all three actions. The degree of the polynomial was again chosen as 10, so that the continuous prior distribution was substituted by 10 nodes.

Also the Γ-maximin action can be found graphically. Figure 3 shows the values of the expected utility of the actions from the example in dependence on σ, which is calculated with the help of discretizations with 10 nodes. Action a_2 has the highest minimal expected utility, so it is Γ-maximin action.

5.2 General Algorithms

The method exemplified here is quite general. In more complex situations, with less smooth utility functions, or when the utility functions are very similar to each other, the number of nodes can be enlarged to obtain calculations of sufficient accuracy. This would increase the computational effort, but does not make a substantial difference. Of course, especially when the set of actions is large, graphical solutions may be insufficient, and general algorithms are needed. For that purpose, also discretize Ψ, resulting in a grid $\psi_1, \ldots, \psi_s, \ldots, \psi_Q$ of different values.[13]

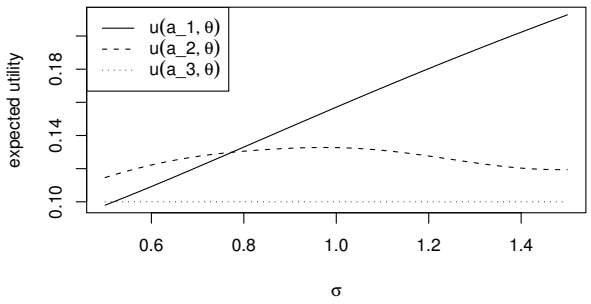

Figure 3: Expected utility depending on σ

That way a *finite* number of probability distributions from $\mathcal{M}$ is processed, each of which is discretized by one of Luceño's theorems, and so eventually a *finite set* of probability distributions with a *finite domain* is considered. In such a setting the algorithm to determine Γ−Maximin solutions described in detail in [2, Theorem 1] can be used mutatis mutandis.[14]

Also an algorithm to determine E-admissible actions can be obtained. For its construction, consider the elements $\pi_{\psi_1}(\cdot), \ldots, \pi_{\psi_Q}(\cdot)$ of $\mathcal{M}$ corresponding to the parameter values $\psi_1, \ldots, \psi_s, \ldots, \psi_Q$. Note that, with defining for all $l = 1, \ldots, n$ and $s = 1, \ldots, Q$,

$$z_{sl} := \mathbb{E}_{\pi_{\psi_s}} u(a^*, \theta) - \mathbb{E}_{\pi_{\psi_s}} u(a_l, \theta), \tag{15}$$

an action a^* is E-admissible, if

$$\exists s \; \forall l : z_{sl} \geq 0\,, \tag{16}$$

or equivalently if

$$\exists s : z_s := \min_{l=1,\ldots,n} a_l \geq 0\,, \tag{17}$$

which is the case iff the optimum $(z_1^*, \ldots, z_Q^*)$ of the following optimization problem

$$\begin{aligned} \sum_{s=1}^{Q} z_s \quad &\rightarrow \quad \max \\ z_{jl} \quad &\geq \quad z_j \quad \forall l = 1, \ldots, n, \;\; \forall j = 1, \ldots, Q, \end{aligned}$$

[13] If the parametrization of the elements of $\mathcal{M}$ is continuous, as is the case in the commonly used statistical models, no substantial loss of information should occur as long as Q is sufficiently large.

[14] Only the set $\mathcal{E}(\mathcal{M})$ arising there in Equation (16) has to be replaced by $\mathcal{Q}$, and the states of natures have to be redefined appropriately.

has a component $z^*_{s_0}$ which is non-negative. The fact that the expectations in (15) can be approximated according to (11) yields directly an algorithm based on linear optimization.

6 Application of Type-II discretization

6.1 Construction of the Discretized Prior Information

Now we turn to Type-II discretization, which produces finally an interval probability on a *finite* state of natures Θ_d with corresponding structure $\mathcal{P}$. In essence, we are now directly in a situation where we can apply algorithms from [9] and [17] to determine the E-admissible actions and the Γ-maximin action(s). To illustrate the general procedure we discuss in some detail the case where $\mathcal{M}$ is a set of parametric distributions, just as before, but now $\mathcal{M}$ is understood as the prestructure of an interval probability, i.e. our prior information consists of the lower and upper envelopes of $\mathcal{M}$, and therefore we explicitly take, for instance, also convex mixtures of elements of $\mathcal{M}$ into consideration allowing for some ambiguity in the shape of the distributions.

Firstly, in order to define the nodes, a reference distribution $\pi_0(\cdot) \in \mathcal{M}$ is chosen, which should be located in the "middle" of $\mathcal{M}$; in its neighborhood there are all possible prior distributions. As described in Section 4 this reference distribution is discretized with N nodes $x_1, \ldots, x_N$, and based on this the new space Θ_d with the elements $\varphi_1, \ldots, \varphi_N$ is obtained.[15] When constructing the interval probability $P(\cdot)$ on $(\Theta_d, \mathcal{P}o(\Theta))$, the next step is to determine the infima and suprema in (12) and (13) from $\mathcal{M}$.[16] For this purpose, for every element $\pi(\cdot) \in \mathcal{M}$ the distribution on the discretized space Θ_d has to be determined. Taking the lower envelope over all these distributions, we confine ourselves for complexity reasons to a support consisting of connected intervals (and then apply natural extension). This means we take for every $\pi(\cdot) \in \mathcal{M}$[17] the lower and upper envelope, resulting in

$$\underline{b}_{s,t} := L\Big(\bigcup_{l=s}^{t-1}\{\vartheta_l\}\Big) \quad \text{and} \quad \overline{b}_{s,t} := U\Big(\bigcup_{l=s}^{t-1}\{\vartheta_l\}\Big). \quad (18)$$

An essential building block in the whole discretization procedure are the nodes obtained by discretizing the reference distribution: Its weights are in this context much less important than the location of the nodes. The location determines the intervals ϑ_j, constituting the states, and so finally the bounds $\underline{b}_{s,t}$ and $\overline{b}_{s,t}$. So a discrete variable is aimed at, the distribution of which approximates the continuous distribution function as exactly as possible. This means, Proposition 2 should be used here.

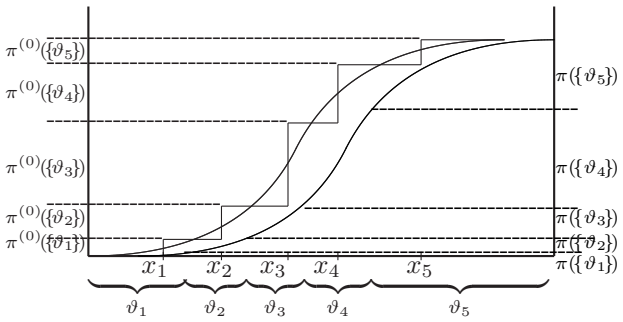

Figure 4: Finding of lower and upper probabilities with the neighborhood of a reference distribution (sketch). For details see footnote 15.

6.2 E-admissibility

If one now considers again the optimization problem determining the expected utility, one has to respect that the considered distributions are not known explicitly, they just have to satisfy the condition

$$L\Big(\bigcup_{l=s}^{t-1}\{\vartheta_l\}\Big) \leq \pi\Big(\bigcup_{l=s}^{t-1}\{\vartheta_l\}\Big) \leq U\Big(\bigcup_{l=s}^{t-1}\{\vartheta_l\}\Big).$$

Next to the auxiliary conditions for λ, i.e., $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0 \quad \forall i$, now therefore also the constraints on $\pi_d(\cdot) \in \mathcal{P}$ have to be considered.

This can be solved by adopting the algorithm developed independently by ([9]) and ([17]): With the help of linear optimization it is possible to decide in the situation with a discrete, but ambiguous state distribution whether an action a_i is E-admissible or not. For this purpose, for every action a_i, the set of all probability measures from the structure $\mathcal{P}$, for which

[15] It may be helpful to look at Figure 4, which sketches graphically several steps of the described procedure to find the lower and upper bounds. The reference distribution chosen is depicted here together with one other distribution from $\mathcal{M}$.
The curve in the middle shows the reference distribution π_0, while the steps represent its discretization with nodes $x_{j,\pi_0}, j = 1 \ldots 5$. The corresponding states ϑ_j are denoted at the abscissa; the probability masses can be seen on the left. As an example, one other distribution is discretized, the corresponding values of the $\pi(\{\vartheta_j\})$ can be read at the right side.

[16] Note that when directly an F-probability on Θ is given, for instance, by one of the neighborhood models used in robust statistics, this step can be skipped. Moreover the methods seem quite attractive to provide a further discretization when p-boxes [5] are considered.

[17] In practical calculations most often a further discretization by considering a grid analogous to 5.2 has to be used.

a_i is optimal, is considered:

$$\Pi_i = \Big\{\pi_d(\cdot) \in \mathcal{P} \Big| \sum_{j=1}^{N} u(a_i, \vartheta_j)\pi_d(\{\vartheta_j\}) \geq \sum_{j=1}^{N} u(a_l, \vartheta_j)\pi_d(\{\vartheta_j\}), \quad \forall l = 1, \ldots, n\Big\}$$

If Π_i is not empty, then there is a classical probability measure in $\mathcal{P}$ under which a_i is optimal and consequently a_i is an E-admissible action.

6.3 Γ-maximin Criterion

Linear programming can be used also for the Γ-maximin criterion. A straightforward, but inefficient possibility is, to find the Γ-maximin action with the help of $n = |\mathbb{A}|$ linear programming procedures, where the expected utility of each action is minimized, and then the action with the highest minimal value has to be found. But it is also possible (and more efficient) to find the optimal action, by considering a single optimization problem. As described in ([17]) the optimization problem:

$$\min_{\pi_d \in \mathcal{P}} \sum_{i=1}^{n} \left(\sum_{j=1}^{N} u(a_i, \theta_j)\pi_d(\{\vartheta_j\}) \right) \lambda_i \longrightarrow \max_{\lambda},$$

subject to the additional constraint $\sum_i \lambda_i = 1$, can be transformed into a single linear programming problem, either by introducing the vertices of the corresponding structure $\mathcal{P}$ or by dualization.[18] Straightforward implementations of the method with dualization and the algorithm described before for the lower and upper bounds have been used in the example below.

6.4 Numerical Example

In the following these algorithms are applied to a numerical example. Let the utility functions $u(a_i, \vartheta)$ of the actions $a_1, \ldots, a_5$ have the form:

$$\begin{aligned} u(a_1, \vartheta) &= 1 \\ u(a_2, \vartheta) &= -(\vartheta - 0.5)^2 + 2.3 \\ u(a_3, \vartheta) &= -(\vartheta + 0.75)^2 + 4.5 \\ u(a_4, \vartheta) &= -|\vartheta - 1| + 2.1 \qquad \text{and} \\ u(a_5, \vartheta) &= -\frac{(\vartheta - 1)^2}{4} + 1.5. \end{aligned}$$

Again we assume that $\mathcal{M}$ consists of all normal distributions with $\mu \in [0.75, 1.25]$ and $\sigma \in [0.75, 1.25]$, but, as discussed above, we explicitly want to allow for ambiguity concerning the type of the distribution and therefore take $\mathcal{M}$ only as a prestructure (cf. [21]), i.e. as a building block to construct an interval-valued assignment - and a corresponding structure (set of compatible distributions) - by passing over to the lower and upper envelope. Firstly the lower and upper bounds have to be found. A normal distribution with $\mu = 1$ and $\sigma = 1$ seems to be a natural choice for the reference distribution. This distribution is now discretized with a fixed number N of nodes. Accordingly normal distributions, which are inside the given bounds for μ and σ^2 are used to find the interval limits $\underline{b}_{s,t}$ and $\overline{b}_{s,t}$ in (18). In the first part of this example the discretization method based on Proposition 2 with $N = 3$ and $M = 10$, i.e. together 30 nodes, has been chosen. To find the bounds, the normal distributions with $\mu \in [0.75, 0.76, \ldots, 1.25]$ and $\sigma \in [0.75, 0.76, \ldots, 1.25]$ have been considered. Implementation of the algorithms from ([17]) yields for the criterion of E-admissibility the vector $(0, 1, 1, 1, 1)$: actions a_2, a_3, a_4 and a_5 are E-admissible under these constraints, action a_1 is not E-admissible. The optimal action under the Γ-maximin criterion is a_5 ($\lambda = (0, 0, 0, 0, 1)$) with a minimal expected utility of 1.106. For comparison, the same calculation has been made with the discretization method based on Proposition 1 with $N = 30$ nodes. The resulting values differ: just a_4 and a_5 are E-admissible actions, while a_5 is also here Γ-maximin action with a minimal expected utility of 1.087.

6.5 Notes on the Accuracy of the Results

It is certainly better to use the method of Proposition 2: it produces a new random variable with a distribution function which is more similar to the shape of the original distribution than the function of a variable produced with the ordinary Gaussian quadrature rule. Indeed the results, as seen above, are different. For explanation see the following Figure 5. It shows the differences between the application of both propositions and their consequences for finding the lower and upper bounds: The new variable produced with the method in Proposition 1 shows big differences to the distribution function of the original distribution, while the curve of the second method can hardly be distinguished from the continuous distribution (picture on the left). The number of nodes was in the first theorem $N = 60$, while in the second one with $N = 3$ and $M = 20$ was used, leading altogether again to 60.

The relatively bad approximation of the original CDF by Proposition 1 follows from the fact that in the simple Gaussian quadrature the nodes for the whole sup-

[18] A second advantage of this algorithm is the fact, that it considers also the mixed extension of the set of actions: the Γ-maximin action does not necessarily have to be a pure i.e. non-randomized action (see [2]).

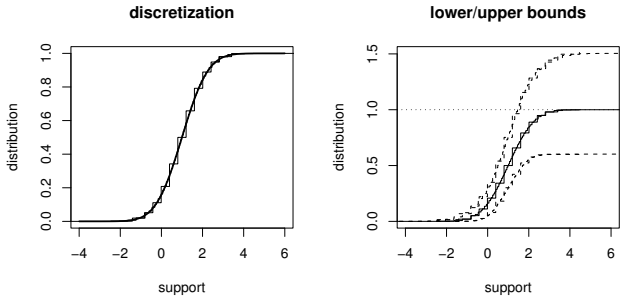

Figure 5: Discretization with Proposition 1 respectively 2 (left). Lower/upper bounds for the probabilities $\pi(\{\vartheta_j\})$ calculated with both methods (right).

port are chosen with one single polynomial. For this reason a lot of nodes are located in the less interesting outer areas of the support. And, as explained above, the locations of the nodes are more important to find the bounds $\underline{b}_{s,t}$ and $\overline{b}_{s,t}$ than the weights (which would be no problem for the method in Proposition 1). The approximation in the inner areas of the support is much exacter with the second method: with a clever choice of the intervals there are a lot of nodes in the important areas. At the end both methods lead to different values of $\overline{b}_{s,t}$'s and $\underline{b}_{s,t}$'s, whereas the values obtained by applying Proposition 2 are to be preferred as explained above. On the right hand side of Figure 5 one can see these differences in the displayed $\underline{b}_{j-1,j}$ and $\overline{b}_{j-1,j}$. The curves in the middle are the discretizations of the reference distribution, below there are the $\underline{b}_{j-1,j}$, above the $\overline{b}_{j-1,j}$.

For the differences between both methods concerning the evaluation of the Γ-maximin action watch Figure 6 which shows the results of the linear optimization for obtaining the Γ-maximin solution with both methods.

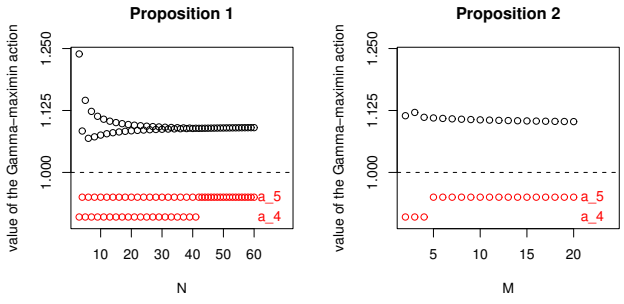

Figure 6: Γ-maximin action, found with Theorem 1 respectively Theorem 2. Upper part of the figure: expected utility value of the Γ-maximin action, lower part: which action is Γ-maximin?

In the upper part of each graph one can see the calculated minimal expected utility value of the Γ-maximin action. The lower part shows the corresponding action. The results based on Proposition 1 oscillate at the beginning between a_5 and a_4 and stay finally stable at action a_5. Also with the method of Proposition 2 (with $N = 3$) the Γ-maximin action at the beginning is a_4. But relatively quickly, from $M = 4$ on, which is discretization with 12 nodes, a_5 stays the optimal solution.

The expected values oscillate with both methods in the same way. But with a high number of nodes the results become more stable. As explained above, the results of Proposition 2 are better in a case, where the results differ.

7 Concluding Remarks

We have discussed a sophisticated method for discretizing a continuous random variable. In contrast to straightforward ad-hoc discretizations, for instance by rounding, one is able to enforce important relations between both variables: the discrete variable and the continuous variable have a given number of moments in common, and also their distribution functions can be ensured to coincide in a certain set of points.

In our view this makes the method quite attractive in imprecise probability theory far beyond decision theory, where we have exemplified the power of the method in two typical applications. Further fruitful areas of application include the calculation of posterior probabilities from the generalized Bayes rule for continuous distributions and the extension of graphical models based on continuous distributions to imprecise probabilities.

Of course, the presentation given here is mainly an exploratory sketch of some basic ideas. Deeper investigations are urgently needed in order to find general recommendations on the trade-off between complexity and accuracy of the approximation. In this respect, also special attention has to be paid to the utility function, in particular when it is not smooth.[19] Another important detail is the sensitivity of the results with respect to the choice of the reference distribution. In rare case only, like the application in neighborhood models (see the survey in [3, Section 4] as well as [7], [20], [14]), there is a unambiguously natural candidate, and canonical examples providing well-accepted recommendations have still to be developed.[20]

[19] One referee suggested to utilize duality of utility and probability for this purpose and to take discreteness of the utility function explicitly into account as well.

[20] One general idea in this direction we owe a referee, who suggested to choose that distribution in $\mathcal{M}$ which minimizes

8 Acknowledgement

We wish to thank all three referees for their helpful and very stimulating remarks.

References

[1] A. S. Ahmad. *Statistical Analysis of Heaping and Rounding Effects.* Dissertation. Department of Statistics, Univ. of Munich, 2006.

[2] T. Augustin. Expected Utility within a Generalized Concept of Probability - a Comprehensive Framework for Decision Making under Ambiguity. *Stat. Pap.*, 43: 5–22, 2002a.

[3] T. Augustin. Neyman-Pearson Testing under Interval Probability by Globally Least Favorable Pairs. Reviewing Huber-Strassen Theory and Extending it to General Interval Probability. *J. Stat. Plan. Inf.*, 105: 149–173, 2002b.

[4] F. G. Cozman. Credal Networks. *Artif. Intell.*. 120: 199–233, 2000.

[5] S. Ferson et al. *Constructing probability boxes and Dempster-Shafer structures.* Technical Report (SAND 2002-4015), Sandia National Laboratories, 2003.

[6] W. Gautschi. Algorithm 726: ORTHPOL – A Package of Routines for Generating Orthogonal Polynomials and Gauss-Type Quadrature Rules. *ACM Trans. Math. Software*, 20: 21–62, 1994.

[7] P. J. Huber and V. Strassen. Minimax Tests and the Neyman-Pearson Lemma for Capacities. *Ann. Stat.*, 1: 251–263, Correction, 2: 223–224, 1973.

[8] J. F. Kenney and E. S. Keeping. *Mathematics of Statistics, Part 2*, 2nd edition, D. Van Nostrand Company, INC, 1951.

[9] D. Kikuti, F. G. Cozman and C. P. de Campos. Partially Ordered Preferences in Decision Trees: Computing Strategies with Imprecision in Probabilities, In: R. Brafman, U. Junker (eds.): Multidisciplinary IJCAI-05 Workshop on Advances in Preference Handling, Edinburgh, Scotland, 118–123, 2005. see also `http://wikix.ilog.fr/wiki/pub/Preference05/WsProceedings/Pref05.pdf`

[10] A. Luceño. Discrete Approximations to Continous Univariate Distributions – an Alternative to Simulation. *J. Roy. Stat. Soc.*, B 61: 345–352, 1999.

[11] M. Obermeier. *Luceños Diskretisierungstechnik für stetige Verteilungen — Implementierung und Anwendung.* Diploma thesis, Dep. of Statistics, Univ. of Munich, 2006.

[12] R.A. Sack and A.F. Donavan. An Algorithm for Gaussian Quadrature given Modified Moments. *Numer. Math.*, 18: 465–478, 1971/72.

[13] W. F. Sheppard. On the Calculation of the most Probable Values of Frequency Constants for Data Arranged according to Equidistant Division of a Scale. *Proc. London Math. Soc.*, 29: 353–380, 1898.

[14] D. Skulj. Jeffrey's Conditioning Rule in Neighbourhood Models. *Int. J. Approx. Reasoning*, 42: 192–211, 2006.

[15] J. Stoer and R. Bulirsch. *Numerische Mathematik I.* Springer, 1999.

[16] M. C. M. Troffaes. Decision Making under Uncertainty using Imprecise Probabilities. *Int. J. Approx. Reasoning*, in press, 2007.

[17] L. V. Utkin and T. Augustin. Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In: F.G. Cozman, R. Nau and T. Seidenfeld (eds.) *Proc. of the 4th Int. Symposium on Imprecise Probabilities and their Applications* (ISIPTA '05), 349–358, 2005.

[18] S. B. Vardeman. Sheppard's Correction for Variances and the "Quantization Noise Model". *IEEE Trans. Instr. Meas.*, 54: 2117–2119, 2005.

[19] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, 1991.

[20] A. Wallner. Bi-elastic Neighbourhood Models. In: J.-M. Bernard, T. Seidenfeld, and M. Zaffalon (eds) *Proc. of the 3rd Int. Symposium on Imprecise Probabilities and their Applications (ISIPTA '03)*, 593–607, 2003.

[21] K. Weichselberger. *Elementare Grundbegriffe einer allgemeinen Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept.* Physica, 2001.

[22] J. C. Wheeler. Rocky Mountain. *J. Math.*, 4: 287–296, 1974.

the maximal Kullback-Leibler distance over $\mathcal{M}$.