Decision making under incomplete data using the imprecise Dirichlet model

L.V. Utkin
Department of Computer Science
St.Petersburg State Forest Technical Academy
Russia
lvu@utkin.usr.etu.spb.ru

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

Abstract

The paper presents an efficient solution to decision problems where direct partial information on the distribution of the states of nature is available, either by observations of previous repetitions of the decision problem or by direct expert judgements.
To process this information we use a recent generalization of Walley's imprecise Dirichlet model, allowing us also to handle incomplete observations or imprecise judgements. We derive efficient algorithms and discuss properties of the optimal solutions. In the case of precise data and pure actions we are surprisingly led to a frequency-based variant of the Hodges-Lehmann criterion, which was developed in classical decision theory as a compromise between Bayesian and minimax procedures.

Keywords. Belief functions, coarse data, decision making, Hodges-Lehmann criterion, imperfect measurements, imprecise Dirichlet model, imprecise probabilities, incomplete data, interval probability, interval statistical models, predictive probabilities, set-valued observations, statistical data

1 Introduction

In real decision situations decision makers have only limited information about probability distributions involved in the problem, and decision analysis is associated with large uncertainty. As a result, decision makers are often confronted with the problem that the very demanding and strong conditions of the classical probability calculus, and the decision models based on it, are often not satisfied. In these cases it is natural to ask how to take into account the limitation of information and what conclusions can be drawn on the basis of such limited information.

Various tools for a more general and sophisticated uncertainty representation can be found in the literature, including Dempster-Shafer structures [16, 38], interval-valued probabilities [47], imprecise probabilities [30, 45], etc. The corresponding decision making models have been developed in accordance with the different types of the uncertainty representation (e.g., [1, 3, 18, 33, 21, 34, 39, 43, 49].) These models allow to handle partial information on the stochastic behaviour on the state of nature.
Here we take explicitly into account the construction of the information and consider decision problems where direct data on the states are available. The data are of multinomial structure, consisting of independent categorical observations. As usual, real values can be associated with the observations as long as the ordering in these values is not understood as providing additional information. In addition, the model seems also to be suitable for processing expert judgements, as long as they are based on independent sources of information.

A particular attractive feature of our method is that it will prove able to incorporate even set-valued observations, i.e., to handle situations where the corresponding category can not be observed exactly and is only known to belong to a certain subset of the sample space. This is an important issue in many applications, but up to now there is no unique terminology. Depending on the context, different terms are common, like 'coarse data', or 'incomplete data', to denote such data sets as a whole, and 'imperfect measurement' or 'interval-valued observations', to denote the single set-valued observations.

To process (complete) multinomial data a Bayesian would recommend to use the Dirichlet model (for ease of distinction called *precise Dirichlet model* (PDM) in the sequel). The PDM has been widely adapted to many applications due to interesting statistical properties, in particular, due to the important fact that the Dirichlet density functions constitute a conjugate family of density functions with respect to multinomial likelihoods. A very promising generalization of the PDM, taking into account lack of prior information,

is Walley's *imprecise Dirichlet model* (IDM), (cf. [46]; for a recent survey of applications see [7]).

This paper applies the IDM to decision making and derives simple algorithms for computing optimal randomized and pure actions. The method developed solves two practically important problems that can not be addressed by any of the classical approaches to decision theory in a satisfying manner: First of all, relying on the IDM enables us to take into account explicitly that the number of judgements or measurements may be rather small, i.e. much too small for being able to apply asymptotic arguments, based on the consistent estimation of the distribution of the states of nature. Secondly, we allow information about states of nature to be represented by imprecise, i.e, for instance, interval-valued observations or measurements. It turns out that this general case can be handled by considering a set of IDMs, which interestingly leads to some extension of the decision problem [34, 39, 49, 21] using Dempster-Shafer theory for uncertainty representation.

The paper is organized as follows: In Section 2 we formulate the problem under consideration more precisely. After having recalled some basic aspects of the imprecise Dirichlet model in Section 3, we apply it in Section 4 to the decision problem. In Section 5 we derive algorithms to determine the optimal randomized and unrandomized actions under a pessimistic criterion relying of strict ambiguity aversion. Section 6 extends consideration to imprecise observations and judgements. Close relations to Dempster-Shafer decision making will be illuminated, and a numerical example will be analyzed. Section 7 glances at more complex decision criteria and Section 8 concludes with some final remarks.

2 Statement of the decision problem

Consider the basic model of decision theory: One has to choose an *action* from a non-empty, finite set $\mathbb{A} = \{a_1, \ldots, a_r, \ldots, a_n\}$ of possible actions. The consequences of every action depend on the true, but unknown *state of nature* $\omega \in \Omega = \{\omega_1, \ldots, \omega_j, \ldots, \omega_m\}$. The corresponding outcome is evaluated by the *utility function* $u : (\mathbb{A} \times \Omega) \to \mathbb{R}$, $(a, \omega) \longmapsto u(a, \omega)$ and by the associated random variable $\mathbf{u}(a)$ on $(\Omega, \mathcal{P}o(\Omega))$ taking the values $u(a, \omega)$.[1] Often it makes sense to study *randomized actions*, which can be understood as a probability measure $\lambda = (\lambda_1, \ldots, \lambda_r, \ldots, \lambda_n)$ on $(\mathbb{A}, \mathcal{P}o(\mathbb{A}))$. Then $u(\cdot)$ and $\mathbf{u}(\cdot)$ are extended to randomized actions by defining $u(\lambda, \omega) := \sum_{r=1}^{n} u(a_r, \omega)\lambda_r$.

[1] Alternatively a loss function $l(a, \omega)$ is assigned, which can be embedded into the framework proposed by setting $u(a, \omega) = -l(a, \omega)$.

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 (e.g. an ideal lottery), and the corresponding probability mass function $\pi(\cdot)$ on the sample space Ω is completely known, then the Bernoulli principle is nearly unanimously favored. One chooses that action which maximizes the expected utility

$$\mathbb{E}_\pi \mathbf{u}(a_r) := \mathbb{E}_\pi \mathbf{u}_r = \sum_{j=1}^{m} u(a_r, \omega_j) \cdot \pi(\omega_j) \qquad (1)$$

among all $r = 1, ..., n$. For simplicity, the obvious constraints $\lambda_j \geq 0$ will be omitted in most places, and we denote

$$\pi_j := \pi(\omega_j),\ u_{rj} := u(a_r, \omega_j), \underline{\mathbf{u}}_r := \min_{j=1,\ldots,m} u_{rj},$$
$$\overline{\mathbf{u}}_r := \max_{j=1,\ldots,m} u_{rj},\ \mathbf{1} := (1, ..., 1)^{\mathrm{T}}.$$

.

Here we aim at developing a powerful method to solve such decision problems when $\pi(\cdot)$ is not known, but – potentially imprecise – data from previous repetitions or expert judgements are available. If one had infinitely many – precise – observations, one would be able to apply the Bernoulli principle based on a consistent plug-in estimator of $\pi(\cdot)$. In applications this is rarely feasible and alternative methods are needed taking into account the lack of complete information explicitly. Therefore, we base our proposal on Walley's IDM [46], and a recent generalization of it [42].

3 Walley's imprecise Dirichlet model

The observations ω_j are assumed to be categorical and unordered, and to be independently and identically distributed. Then the corresponding multivariate random quantity counting the number of occurrences of the j-th category possesses a multinomial distribution with parameter vector π. In a Bayesian setting the corresponding conjugated prior is the (precise) Dirichlet $(s, \mathbf{t})$ prior distribution (e.g. [15]), where $\mathbf{t} = (t_1, ..., t_m)$ is a certain element of the interior of the unit simplex denoted by $S(1, m)$.The parameter $t_i \in (0, 1)$, $i = 1, \ldots, m$ is the mean of π_i under the Dirichlet prior; the hyperparameter $s > 0$ determines the influence of the prior distribution on posterior probabilities. When multiplied with multinomial likelihood function, the Dirichlet $(s, \mathbf{t})$ prior density generates the Dirichlet $(N + s, \mathbf{t}^*)$ density with $t_j^* = (n_j + st_j)/(N + s)$.

An important argument against the use of the PDM is that – at least without a huge amount of observations

– inferences depend on the value of $\mathbf{t}$ to be fixed in advance, typically without having sufficient information to guide the choice. Moreover, there is the desire for a model where the predictive probabilities used in decision making directly reflect the sample size, i.e. the amount of statistical information available.

Both problems are solved by the imprecise Dirichlet model as defined by Walley [46]. It avoids unjustifiable prior assumption by relying on the set of *all* Dirichlet $(s, \mathbf{t})$ distributions such that $\mathbf{t} \in S(1, m)$, and the inferences depend – via the width of the intervals for the predictive probability obtained – on the sample size. In the IDM, there is a *hyperparameter* s determining how quickly upper and lower probabilities of events converge as statistical data accumulate. s can be interpreted as the number of observations needed to reduce the imprecision (i.e., the difference between upper and lower probabilities) to half its initial value, or as the notional number of yet unseen observations. Consequently, smaller values of s produce faster convergence and stronger conclusions, whereas large values of s produce more cautious inferences. At the same time, the value of s must not depend on m or the number of observations. A detailed discussion concerning the parameter s and the IDM can be found in particular in [7, 46].

Let A be any non-trivial subset of the sample space $\Omega = \{\omega_1, ..., \omega_m\}$, i.e., A is not empty and $A \neq \Omega$, and let $n(A)$ denote the observed number of occurrences of A in N trials, $n(A) = \sum_{\omega_j \in A} n_j$ where $n_j := n(\{\omega_j\})$. The predictive probability $P(A|\mathbf{n}, \mathbf{t}, s)$ under a certain Dirichlet posterior distribution is

$$P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A) + st(A)}{N + s},$$

where $t(A) := \sum_{\omega_j \in A} t_j$. It should be added that $P(A|\mathbf{n}, \mathbf{t}, s) = 0$ if A is empty, and $P(A|\mathbf{n}, \mathbf{t}, s) = 1$ if $A = \Omega$. By maximizing and minimizing $P(A|\mathbf{n}, \mathbf{t}, s)$ over $\mathbf{t} \in S(1, m)$, Walley [46] obtains the posterior lower and upper predictive probabilities of A as:

$$\underline{P}(A|\mathbf{n}, s) = \frac{n(A)}{N + s}, \qquad \overline{P}(A|\mathbf{n}, s) = \frac{n(A) + s}{N + s}.$$

4 Decision making by using the imprecise Dirichlet model

As a preparation, let us briefly consider an approach for decision making under the condition that π obeys to the PDM and there are perfect observations of states of nature, namely the numbers $(n_1, ..., n_m)$ of occurrences of $\omega_1, ..., \omega_m$ in the N trials. By relying on Bayesian methodology, the expected utility of an action λ is calculated as follows:

$$\mathbb{E}\mathbf{u}(\lambda) = \int_{S(1,m)} \sum_{i=1}^{m} (u(\lambda, \omega_i) \cdot \pi_i)\, p(\pi) \mathrm{d}\pi$$

$$= \sum_{i=1}^{m} u(\lambda, \omega_i) \cdot \int_{S(1,m)} \pi_i p(\pi) \mathrm{d}\pi = \sum_{i=1}^{m} u(\lambda, \omega_i) \cdot \mathbb{E}_p \pi_i,$$

where $\mathbb{E}_p \pi_i = \frac{n_i + st_i}{N+s}$,finally resulting in

$$\mathbb{E}\mathbf{u}(\lambda) = \sum_{i=1}^{m} u(\lambda, \omega_i) \frac{n_i + st_i}{N + s}. \tag{2}$$

Passing over to the IDM leads to lower and upper expected utilities arising from the following optimization problems:

$$\underline{\mathbb{E}}\mathbf{u}(\lambda) = \inf_{\mathbf{t} \in S(1,m)} \mathbb{E}\mathbf{u}(\lambda), \qquad \overline{\mathbb{E}}\mathbf{u}(\lambda) = \sup_{\mathbf{t} \in S(1,m)} \mathbb{E}\mathbf{u}(\lambda).$$

In literature several criteria have been suggested to compare the interval-valued expected utility

$$[\underline{\mathbb{E}}\mathbf{u}(\lambda), \overline{\mathbb{E}}\mathbf{u}(\lambda)] \tag{3}$$

of actions λ (see, in particular, the recent survey by [41] and [44], who give further references).[2] The first branch, like the criterion of maximality (as proposed by [45]) or the concept of E-admissibility (advocated by [31], [35]), renounces the completeness of the ordering and generalizes the concept of admissibility by distinguishing a set of actions as being not inferior.

On the other hand, often a complete ordering of the actions is desired, and the interval-valued expected utility eventually has to be transformed to the real line. The most conservative choice is to be strictly ambiguity averse, concentrating on the lower interval limit only. (Section 5 and Section 6.1 will rely on this criterion, while Section 7 briefly will consider alternative criteria.)

5 Decision making under strict ambiguity aversion

Under strict ambiguity aversion an action is optimal iff for all λ the inequality $\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geq \underline{\mathbb{E}}\mathbf{u}(\lambda)$ is satisfied.[3] An algorithm is described in

[2] For rigorous axiomatic justification of generalized expected utility in the sense of (3) and different criteria derived from it see among others [19] and [20], as well as the references therein, who manage to extend Neumann-Morgenstern and Anscombe-Aumann theory to the situation of complex uncertainty with partial prior information.

[3] This criterion has been proposed under different names. It corresponds to the Gamma-Minimax criterion (as considered, e.g., in [5, Section 4.7.6]), to the Maxmin expected utility model ([22]), to the MaxEMin criterion investigated by [28] (cf. also [29] and the references therein) and the notion of maximinity in [45]. In the case of two-monotone capacities it is equivalent to maximizing Choquet expected utility (as studied, e.g., in [9]).

Proposition 1 *The optimal randomized action λ^* satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geq \underline{\mathbb{E}}\mathbf{u}(\lambda)$ for all λ is determined by solving the following linear programming problem:*

$$\max_{\lambda, G} G \tag{4}$$

subject to $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$, and for $j = 1, ..., m$,

$$G \leq \frac{1}{N+s} \sum_{r=1}^{n} \lambda_r \left(s \cdot u_{rj} + \sum_{i=1}^{m} u_{ri} \cdot n_i \right). \tag{5}$$

Proof. It follows from (2) that λ^* is found by considering

$$\inf_{\mathbf{t} \in S(1,m)} \sum_{i=1}^{m} \sum_{r=1}^{n} u_{ri} \lambda_r \cdot \frac{n_i + st_i}{N+s} \to \max_{\lambda}. \tag{6}$$

subject to $\lambda \cdot \mathbf{1} = 1$. For solving this problem, let us adapt [1, 3], who suggested to introduce a new variable $G = \inf_{\mathbf{t} \in S(1,m)} \mathbb{E}\mathbf{u}(\lambda)$. Then problem (6) is equivalent to a problem with objective function (4) and with the infinite number of constraints

$$G \leq \sum_{i=1}^{m} \sum_{r=1}^{n} u_{ri} \lambda_r \cdot \frac{n_i + st_i}{N+s}, \quad \mathbf{t} \in S(1,m), \tag{7}$$

and $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$. Following [1, 3] further, note that the constraints are already satisfied, if they are satisfied for all extreme points of the convex sets of distributions defined by $\underline{P}(\cdot|\mathbf{n}, s)$ and $\overline{P}(\cdot|\mathbf{n}, s)$, which are simply obtained by considering the extreme points of $S(1,m)$. The latter have the form $(1, 0, ..., 0)$, $(0, 1, ..., 0), ...,$ $(0, 0, ..., 1)$. Therefore, (7) is reduced to the set of m linear constraints described in (5), as was to be proven.[4] ■

Restricting λ_r to be either 0 or 1 optimal unrandomized actions are obtained via

Corollary 1 *The optimal unrandomized action (pure action) a^* satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}_r \geq \underline{\mathbb{E}}\mathbf{u}_k$ for all $k = 1, ..., n$ is determined by solving the following problem:*

$$\frac{1}{N+s} \left(\sum_{i=1}^{m} u_{ri} n_i + s \cdot \underline{\mathbf{u}}_r \right) \to \max_{r=1,...,n}. \tag{8}$$

It can be seen from (8) that the objective function is nothing else but a mixture of two criteria: the criterion of maximum expected utility, with probabilities $\pi(\{\omega_i\})$ replaced by the corresponding relative frequency n_i/N, and Wald's criterion. The weights $N/(N+s)$ and $s/(N+s)$, respectively, are directly connected to the uncertainty involved, depending on the hyperparameter s and the sample size N. Consequently, when $N = 0$, i.e., before any observation on the states of nature is available, Wald's criterion is used. On the other hand, when the sample size tends to infinity, i.e. when there are enough data to estimate $\pi(\cdot)$ consistently, expected utility based on the estimate of $\pi(\cdot)$ is used. In this sense, (8) is a frequency-based type of Hodges-Lehmann criterion (cp. [26]), which has been proposed in classical decision theory as a compromise between the Bayesian and the minimax approach. — Equation (8) also provides a behaviourial interpretation of the hyperparameter s: In principle, one could develop canonical examples to determine a decision maker's value of s in an experimental manner.

A closer investigation of (8) shows, however, also a possibly unwanted effect. When the utility function is such that $\underline{u}_r$ is the same for all actions a_r then the second summand in (8) does not matter and $\frac{\sum u_{ri} \cdot n_i}{N+s}$ is maximized, which distinguishes the same action as optimal as the naive frequentist approach where in (6) the probability $\pi(\cdot)$ is replaced by the vector of observed relative frequencies. – Note that, in principle, such a situation can always be constructed, by adding a "bad" state of nature ω_0 (like "crash of the economic system" in the example in Section 6.4) which has constant utility u_0 for all actions such that $u_0 < \min_{r,i} u_{ri}$. This means that – just as in minimax theory – special attention has to be paid to careful selection of the states of nature that are taken into consideration. To turn it into other – even more trenchant – words: in decision theory the IDM's property of satisfying the embedding principle is lost.

6 Decision making under incomplete data

Now consideration is extended to the practically quite important case of imprecise observations, where observations may be too vague to be associated with a certain singleton $\{\omega_j\}$. Instead it is only known that the realized state of nature lies in some subset $A_i \subseteq \Omega$, see, for instance, [25] for examples in the biometric context. Heitjan and Rubin, who have coined the term coarsened for such data, derive in [24] – rather severe – conditions under which the coarsening may be ignored. Blumenthal (cf. [8]) discusses a multinomial model under the additional assumption that the probability distribution of the coarsening process is known.

However, quite often Heitjan and Rubin's so-called coarsening at random assumption is violated and straightforward analysis may be heavily biased. The

[4] An alternative proof can rely on total monotonicity of the lower interval limit derived from the IDM and use (10), again circumventing minima by auxiliary variables.

same applies, as the simulations in [32] show, when the typically unknown distribution in Blumenthal's model is misspecified. Therefore a thorough analysis without relying on unjustified assumptions is highly desirable. Several authors, among them [40], have understood Dempster-Shafer belief functions / random sets as an appropriate tool to model such situations: they use empirical belief functions (see below) based on relative frequencies of the observed subsets A_i for the analysis. Although then imprecision in the observations is taken into account, still a severe bias may occur, because this way to proceed neglects – by implicitly equating relative observed frequencies and probabilities – finite sample variation, which may have, as argued above and in the end, a strong distorting effect, too.

6.1 Extended empirical belief functions

In order to handle both sources of potential bias – imprecision in the observations as well as the limited sample size – we rely on a model recently developed by [42], which, in essence, considers all multinomial models compatible with the data and will lead to a powerful extension of empirical belief function.

To be a bit more precise (for a detailed account see [42]): Data consist now of c_i observations of the subset $A_i \subseteq \Omega$, $i = 1, ..., M$, such that $\sum_{i=1}^{M} c_i = N$. Furthermore, it is helpful to introduce sets J_i denoting the set of indices of states of nature belonging to A_i, i.e. $A_i = \{\omega_j : j \in i\}$. Evidently the standard case is included: there all available subsets consist of single elements, i.e., $A_i = \{\omega_j\}$ and $J_i = \{j\}$.

Translating the situation into an urn model, we have m urns $\omega_1, \ldots, \omega_m$ (corresponding to the states of nature), where ω_i consists of ball with number i. Relying on the notation just introduced we randomly choose subsets A_i of urns and take randomly c_i balls from the urns numbered by elements of J_i. Given M, there exist different possible combinations $k = 1, \ldots, K$ of numbers of balls taken from the urns. Denote the k-th possible vector of balls by $\mathbf{n}^{(k)} = (n_1^{(k)}, \ldots, n_m^{(k)})$, where n_l^k, $l = 1, \ldots, m$ is the number of balls with color l drawn in the M choices, and let $\mathbf{c} = (c_1, \ldots, c_n)$. Assuming that the subsets A_i are independently chosen from the set of all subsets of Ω and that the probability of selecting a ball from the j-th urn is π_j, every combination of balls produces one standard multinomial model. A number of possible combinations of balls produce the same number of standard multinomial models. Moreover, we can not prefer one model over another[5].

Since we have a set of vectors $\mathbf{n}^{(k)}$, then even if we know precisely the conditional probabilities $P(A|\mathbf{n}^{(k)})$ for every event A in Ω and every possible vector $\mathbf{n}^{(k)}$, still we can only compute lower and upper probabilities for events A:

$$\underline{P}(A|\mathbf{c}) = \min_k P(A|\mathbf{n}^{(k)}), \quad \overline{P}(A|\mathbf{c}) = \max_k P(A|\mathbf{n}^{(k)}).$$

As the vectors $\mathbf{n}^{(k)}$ depend on $\mathbf{c}$, the resulting lower and upper probabilities (after minimizing and maximizing $P(A|\mathbf{n}^{(k)})$) depend on $\mathbf{c}$, and so they are denoted by $\underline{P}(A|\mathbf{c})$ and $\overline{P}(A|\mathbf{c})$.

In the case of multinomial samples, the PDM as a prior distribution on the probabilities $P(\omega_j) = \pi_j$, $j = 1, ..., m$, is the traditional choice. However, using the IDM allows us to take into account lack of prior information and the possible fact that the number of observations A_i may be rather small. Then the lower and upper probabilities $\underline{P}(A|\mathbf{c}, s)$ and $\overline{P}(A|\mathbf{c}, s)$ of an event A, corresponding to a set of indices $J \subseteq \{1, \ldots, m\}$, are computed as follows:

$$\underline{P}(A|\mathbf{c}, s) = \frac{\min_k n^{(k)}(A) + s \cdot \inf_{\mathbf{t} \in S(1,m)} t(A)}{N + s},$$

$$\overline{P}(A|\mathbf{c}, s) = \frac{\max_k n^{(k)}(A) + s \cdot \sup_{\mathbf{t} \in S(1,m)} t(A)}{N + s}.$$

where $t(A) := \sum_{j \in J} t_j$, $n^{(k)}(A) := \sum_{j \in J} n_j^{(k)}$.

It can be shown that the resulting lower and upper probabilities of A can be obtained from the observations $A_1, ...A_M$ and $c_1, ...c_M$ as follows:

$$\underline{P}(A|\mathbf{c}, s) = \frac{\sum_{i: A_i \subseteq A} c_i}{N + s}, \overline{P}(A|\mathbf{c}, s) = \frac{\sum_{i: A_i \cap A \neq \varnothing} c_i + s}{N + s}. \tag{9}$$

For a closer investigation of these results, it is beneficial to consider them within the framework of Dempster-Shafer theory (e.g. [16, 38]): With [32] we call a *basic probability assignment* $m : \mathcal{P}o(\Omega) \to [0, 1]$, with $m(\varnothing) = 1$, $\sum_{A \in \mathcal{P}o(\Omega)} m(A) = 1$ and the corresponding *belief* $Bel(A)$ and *plausibility* $Pl(A)$ functions with $Bel(A) = \sum_{A_i : A_i \subseteq A} m(A_i)$, $Pl(A) = 1 - Bel(A^c)$ *empirical*, when $m(A_i) = \frac{c_i}{N}$ based on a vector $c_1, ...c_n$ of $\sum_{i=1}^{M} c_i = N$ observations of $A_i \subseteq \Omega, i = 1, ...n$. It is easy to see that the lower and upper probabilities (9) relate to these belief and

[5] It should be noted that the set of possible vectors of balls $\mathbf{n}^{(k)}$ produced by interval-valued observations is very closely related to the set of all possible completions of missing attributes in models of missing data proposed by de Cooman and Zaffalon [13, 50]. Moreover, the set of possible completions of missing attributes coincide with the set of $\mathbf{n}^{(k)}$ when the observations are of a special type.

plausibility functions[6] in the following way:

$$\underline{P}(A|\mathbf{c},s) = \frac{N \cdot Bel(A)}{N+s}\,,\ \overline{P}(A|\mathbf{c},s) = \frac{N \cdot Pl(A)+s}{N+s}.$$

Moreover, $\underline{P}(A|\mathbf{c},s)$ and $\overline{P}(A|\mathbf{c},s)$ are belief and plausibility functions again, namely with the basic probability assignment $m^*(A_i) = c_i/(N+s)$ for every A_i and the additional basic probability assignment $m^*(A_\infty) = s/(N+s)$, i.e., $\underline{P}(A|\mathbf{c},s)$ and $\overline{P}(A|\mathbf{c},s)$ can be obtained as empirical belief and plausibility functions under the condition that there are s additional observations $A_\infty = \Omega$. With $m(A_i) = c_i/N$ representing the standard empirical assignment, we have $m^*(A_i) = m(A_i) \cdot N/(N+s)$, and, for all $A \neq \Omega$,

$$\underline{P}(A|\mathbf{c},s) = \sum_{i:A_i \subseteq A} m^*(A_i),$$
$$\overline{P}(A|\mathbf{c},s) = m^*(A_\infty) + \sum_{i:A_i \cap A \neq \varnothing} m^*(A_i),$$

which we call *extended empirical* belief and plausibility functions.

6.2 Decision making with extended empirical belief functions

Since the extended empirical belief functions arising from imprecise information about states of nature are belief functions, we can use the approach introduced by [39], which directly relies on belief functions based on some basic probability assignment $m(\cdot)$. Then the lower expected utility can be found as

$$\underline{\mathbb{E}}\mathbf{u}(\lambda) = \sum_{A_k \subseteq \Omega} m(A_k) \cdot \min_{\omega_i \in A_k} u(\lambda, \omega_i). \tag{10}$$

In the situation under consideration described by the basic probability assignment $m^*(\cdot)$, this can be rewritten as follows:

$$\underline{\mathbb{E}}\mathbf{u}(\lambda) = \frac{s}{N+s} \cdot \min_{\omega_i \in \Omega} u(\lambda, \omega_i) + \sum_{k=1}^{M} \frac{c_k}{N+s} \cdot \min_{\omega_i \in A_k} u(\lambda, \omega_i).$$

Efficient handling of the resulting decision problem is summarized in

Proposition 2 *If the probabilities of m states of nature are described by the imprecise Dirichlet model with the hyperparameter s and information about the states is represented in the form of c_i subsets $A_i = \{\omega_j : j \in J_i\}$, $i = 1, ..., M$, such that $\sum_{i=1}^{M} c_i = N$, then the optimal randomized action λ^* satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geq \underline{\mathbb{E}}\mathbf{u}(\lambda)$ for all λ is determined by solving the following linear programming problem:*[7]

$$\frac{1}{N+s}\left(s \cdot V_0 + \sum_{k=1}^{M} c_k \cdot V_k\right) \to \max_{\lambda},$$

subject to $V_0, V_i \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$,

$$V_i \leq \sum_{r=1}^{n} \lambda_r u_{rj},\ i = 1, ..., M,\ j \in J_i,$$
$$V_0 \leq \sum_{r=1}^{n} \lambda_r u_{rj},\ \ j = 1, ..., m.$$

Proof. We introduce new variables $V_i = \min_{j \in J_i} u(\lambda, \omega_j)$, $i = 1, ..., M$, and $V_0 = \min_{j=1,...,m} u(\lambda, \omega_j)$, and substitute them into the objective function (6.2). Constraints to the optimization problem are derived from the definition of V_i, $i = 0, ..., M$. ■

The extension of Corollary 1 to the situation under consideration is provided by

Corollary 2 *The optimal unrandomized action (pure action) a^* satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}_r \geq \underline{\mathbb{E}}\mathbf{u}_k$ for all $k = 1, ..., n$ is determined by solving:*

$$\frac{1}{N+s}\left(s \cdot \underline{\mathbf{u}}_r + \sum_{k=1}^{M} c_k \cdot \min_{\omega_i \in A_k} u_{ri}\right) \to \max_{r=1,...,n}. \tag{11}$$

Proof. If $\lambda_r \in \{0, 1\}$, then $V_k = \min_{\omega_i \in A_k} u_{ri}$ and $V_0 = \underline{\mathbf{u}}_r$ due to conditions $s \geq 0$ and $c_k \geq 0$. ■

6.3 Some additional remarks

Before we illustrate our approach in a short example, some additional remarks may be appropriate. Firstly, in Remark 1 and Remark 2, equivalent alternative ways to proceed should be briefly mentioned. After that the advantages of the extended modelling will be demonstrated in some extreme cases.

Remark 1 *The above results could be also obtained without using the argumentation via belief functions. Indeed, the set of multinomial models considered in Subsection 6.1 produces a set of expected utilities $\mathbb{E}^{(k)}\mathbf{u}(\lambda)$ such that the lower expected utility is*

$$\underline{\mathbb{E}}\mathbf{u}(\lambda) = \min_{k} \inf_{\mathbf{t} \in S(1,m)} \mathbb{E}^{(k)}\mathbf{u}(\lambda)$$
$$= \min_{k} \inf_{\mathbf{t} \in S(1,m)} \sum_{i=1}^{m} u(\lambda, \omega_i) \cdot \frac{n_i^{(k)} + st_i}{N+s}.$$

[6] A detailed study of belief functions derived from the imprecise Dirichlet model is presented in [42].

[7] The problem considered in Proposition 2 has $\sum_{i=1}^{M} |J_i| + m + n + 1$ linear constraints and $2m+1$ optimization variables. (Here $|J_i|$ denotes the cardinality of the set J_i.) If all the subsets A_i consist of the single elements ω_i, resulting in $c_i = n_i$, and $M = m$, then we get back, after simple transformations, the special case studied in Section 4.

This way leads to the same results, but it allows us to get several forms of linear programming problems for computing the optimal randomized action λ^*.[8]

Remark 2 *Since the information about states of nature is represented by means of lower $\underline{P}(A|\mathbf{c}, s)$ and upper $\overline{P}(A|\mathbf{c}, s)$ probabilities of all events $A \in \Omega$, the decision problem can also be solved by means of the approach proposed by [1, 3], where the extreme points needed can be directly derived from the corresponding basic probability assignment $m^*(\cdot)$.*

One of the main shortcomings of using standard empirical belief functions in decision making is that they assign zero probabilities to yet unobserved states of nature. If we had an infinite number of observations, then the fact of zero probabilities could be accepted. However, if we have a finite number of observations (and more often quite a small number), then it can be subjected to criticism. To show that this fact may indeed lead to controversial results, consider the following decision problem: $\mathbb{A} = \{a_1, a_2\}$, $\Omega = \{\omega_1, \omega_2\}$, the utility function is $u_{11} = -1000$, $u_{12} = 1$, $u_{21} = u_{22} = 0$. Suppose that there is only one judgment ($M = 1$) such that $A_1 = \{\omega_2\}$. According to the model relying on $s = 0$, we can write $\underline{\mathbb{E}}\mathbf{u}_1 = 1$ and $\underline{\mathbb{E}}\mathbf{u}_2 = 0$. Hence the optimal action is a_1, i.e. under almost complete ignorance, where we naturally intend to search for the pessimistic solution, we make the optimistic decision because the model using standard empirical belief functions presumes that the probability of state ω_1 is 0, and so it acts as if this state could never be observed. However, if we take $s > 0$, say $s = 1$, then $\underline{\mathbb{E}}\mathbf{u}_1 = (-1000 + 1)/2 = -499.5$ and $\underline{\mathbb{E}}\mathbf{u}_2 = 0$. Hence the optimal action is a_2, showing that the IDM provides a way to avoid this shortcoming.

A related, again rather problematic issue of standard empirical belief functions is that the assignments do not depend on the sample size and therefore – the potentially very high – finite sampling variation is neglected: Consider, for instance, two samples taken from the same sample space, one with $n = 2$, the other one with $n = 200000$.[9] If in the first case $n_1 = 1 = n_2$, and in the second one $n_1 = 100000 = n_2$, then the relative frequencies for the states $1, 2$ are $1/2$ each, in both cases, and therefore also the standard belief functions derived from them are the same, not distinguishing these substantially different situations. In contrast, the extended belief function approach takes into account that the second information is build on a much stronger basis: the imprecision in the assignment arising from the second situation is much less.

[8] As another alternative, the resulting lower and upper probabilities can be interpreted as arising from a special type of a generalized basic probability assignment in the sense of [4], constructed from the set of IDMs that are derived from the different possible observations.

[9] To focus on the essential point of the argument we further assume that all observations are precise.

Table 1: Values of the utility function u_{rj}

		States of nature			
Actions		ω_1	ω_2	ω_3	ω_4
bonds	a_1	6	9	9	8
stocks	a_2	12	7	3	-2
deposit	a_3	7	7	7	7

The problem becomes in particular drastic, when we return to a situation with unobserved states of nature and consider again the numerical example given above. If we have l identical estimates of the second state of nature, then

$$\underline{\mathbb{E}}\mathbf{u}_1 = -\frac{1000 \cdot s}{l+s} + \frac{l}{l+s}, \qquad \underline{\mathbb{E}}\mathbf{u}_2 = 0.$$

When $s = 0$ then neither $\underline{\mathbb{E}}\mathbf{u}_1 = 1$ nor $\underline{\mathbb{E}}\mathbf{u}_2 = 0$ depend on l. That is, our decision, a_2 is the same if we have 1 observation (almost complete ignorance) or 10000 identical observations (sufficient statistical data). If we take $s = 1$, then a_2 is only superior when $l > 1000$.

6.4 Numerical example

Consider the following well-known investment decision-making example. The states of nature are the states of economy during one year: growth - ω_1, medium growth - ω_2, no change - ω_3, low - ω_4. The problem is to decide to choose an action from three possible courses of action with the given rates of return as shown in the body of Table 1.

Suppose that three experts, relying on independent sources of information[10], supply the following judgments concerning the states of economy: two experts ($c_1 = 2$) believe that the state of economy will be "growth" or "medium growth" ($A_1 = \{\omega_1, \omega_2\}$), one expert ($c_2 = 1$) supposes that the state of economy will be "no change" or "medium growth" ($A_2 = \{\omega_2, \omega_3\}$). So, $M = 2$, $N = 3$.
Let us find the optimal randomized action maximizing the lower expected utility under condition $s = 1$. By using Proposition 2, we write the following problem:

$$\max_{\lambda \in \mathbb{R}^3_+, V_i \in \mathbb{R}} \frac{1}{3+1} (1 \cdot V_0 + 2 \cdot V_1 + 1 \cdot V_2)$$

[10] The assumption of independence is crucial for direct applications of the multinomial likelihood underlying the IDM. Relaxations of this assumption are currently under investigation.

subject to $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and

$$\begin{array}{ll} V_1 \leq 6\lambda_1 + 12\lambda_2 + 7\lambda_3 & V_0 \leq 6\lambda_1 + 12\lambda_2 + 7\lambda_3 \\ V_1 \leq 9\lambda_1 + 7\lambda_2 + 7\lambda_3 & V_0 \leq 9\lambda_1 + 7\lambda_2 + 7\lambda_3 \\ V_2 \leq 9\lambda_1 + 7\lambda_2 + 7\lambda_3 & V_0 \leq 9\lambda_1 + 3\lambda_2 + 7\lambda_3 \\ V_2 \leq 9\lambda_1 + 3\lambda_2 + 7\lambda_3 & V_0 \leq 8\lambda_1 - 2\lambda_2 + 7\lambda_3 \end{array}$$

Hence $V_0 = 27/4$, $V_1 = 27/4$, $V_2 = 33/4$, $\lambda_1^* = 7/8$, $\lambda_2^* = 1/8$, $\lambda_3^* = 0$. The optimal lower expected utility is 7.125.

Let us find now the optimal unrandomized action. By using (11) we get $\underline{\mathbb{E}}\mathbf{u}_1 = 6.75$, $\underline{\mathbb{E}}\mathbf{u}_2 = 3.75$, $\underline{\mathbb{E}}\mathbf{u}_3 = 7$. This implies that the third action is optimal.

It should be noted that, by taking $s = 0$, the optimal randomized action would be $\lambda_1^* = 5/8$, $\lambda_2^* = 3/8$, $\lambda_3^* = 0$, with an optimal lower expected utility of 7.75. At a first glance one might be tempted to say that the decision based on $s = 0$ would be better than that based on $s > 0$ because the lower expected utility in the case $s = 0$ is greater than the expected utility based on $s = 1$. However, as discussed in the previous subsection, this decision is incautious because it does not take into account the fact that the number of judgments is very small ($N = 3$). The optimal unrandomized action by $s = 0$ is not unique because $\underline{\mathbb{E}}\mathbf{u}_1 = \underline{\mathbb{E}}\mathbf{u}_3$ and $\underline{\mathbb{E}}\mathbf{u}_2 < \underline{\mathbb{E}}\mathbf{u}_1$ in this case.

7 Other optimality criteria

In this paper up to now only one particular – quite pessimistic – optimality criterion has been studied. As a considerable improvement a more complex criterion of decision making based on a linear combination of lower and upper expectations with the so-called *caution* parameter $\eta \in [0, 1]$ (cp., e.g., [17, 36, 47]) can be considered. The caution parameter reflects the degree of ambiguity aversion; the more ambiguity averse the decision maker is, the higher is the influence of the lower interval limit of generalized expected utility. $\eta = 1$ corresponds to strict ambiguity aversion, $\eta = 0$ expresses maximal ambiguity seeking attitudes. Methods for the choice of η are considered in detail by [36]. In this case, a pure action a_k is distinguished as optimal iff for all $r \in \{1, ..., n\}$

$$\eta\underline{\mathbb{E}}\mathbf{u}_k + (1-\eta)\overline{\mathbb{E}}\mathbf{u}_k \geq \eta\underline{\mathbb{E}}\mathbf{u}_r + (1-\eta)\overline{\mathbb{E}}\mathbf{u}_r\,.$$

The lower expected utility $\underline{\mathbb{E}}\mathbf{u}_r$ is computed by means of (11). The upper expected utility $\overline{\mathbb{E}}\mathbf{u}_r$ can be found in the same way

$$\overline{\mathbb{E}}\mathbf{u}_r = \frac{1}{N+s}\left(s \cdot \overline{\mathbf{u}}_r + \sum_{k=1}^{M} c_k \cdot \max_{i \in J_k} u_{ri}\right).$$

Hence, with $H_r := \eta\underline{\mathbb{E}}\mathbf{u}_r + (1-\eta)\overline{\mathbb{E}}\mathbf{u}_r$,

$$H_r = \frac{s}{N+s}\left(\eta\underline{\mathbf{u}}_r + (1-\eta)\overline{\mathbf{u}}_r\right) + \frac{1}{N+s}\left(\sum_{k=1}^{M} c_k \left(\eta \min_{i \in J_k} u_{ri} + (1-\eta)\max_{i \in J_k} u_{ri}\right)\right).$$

It can be seen that the algorithm for computing the optimal action taking into account both lower and upper expected utilities is similar to the approach proposed in Section 2. However, it should be noted that the randomized action can not be found in the same simple way and efficient algorithms for solving this problem are currently investigated (see also Section 4 of [44]).

Of course, also criteria not necessarily producing a linear ordering of the actions deserve attention (cf. Section 4 above). [44, Section 5] proposes a general method to determine actions in the sense of [45], which can – in the light of Remark 2 – also be used here. In this context it may be briefly noted that the remark from the end of Section 5 still applies: Even under E-admissibility and maximality complex criterion like the IDM's property of satisfying the embedding principle is lost, making the careful choice of the set Ω of possible states a delicate task.[11]

8 Concluding remarks

A method for decision making using the IDM under imprecise information has been proposed in this paper. This approach can also be regarded as some extension of a procedure relying on empirical belief functions. The considered special cases and the numerical example have shown that the method is reasonable even in cases when the number of possible imperfect measurements or interval-valued judgments is very small.

Next to the handling of more sophisticated optimality criteria further research should also include a comparison with other attempts to model the processing of multinomial observations under lacking prior knowledge ([10], [48]) and an alternative decision theoretic approach where the sampling information is handled by decision functions, the more as the use of posterior distributions, as underlying the IDM, is no longer self-evidently justified in imprecise probability theory ([2], see also [27], [37], [23] for related issues.)

[11] For an easy counterexample consider parts of the utility table from Section 6.4, based on the two different sample spaces $\Omega_1 := \{\omega_1, \omega_2\}$ and $\Omega_2 := \{\omega_1, \omega_2, \omega_3\}$. With data $c_1 = c_2 = 1$ and $s = 1$ action a_1 is neither E-admissible nor maximal, when Ω_1 is considered, but optimal when Ω_2 is considered. — However, the violation of the embedding principle seems to be milder here: actions that are optimal remain optimal under refinements of the sample space with respect to yet unobserved categories.

Aknowledgement We thank the three referees for valuable comments.

References

[1] T. Augustin. Expected utility within a generalized concept of probability - a comprehensive framework for decision making under ambiguity. *Statistical Papers*, 43:5-22, 2002.

[2] T. Augustin. On the suboptimality of robust Bayesian procedures from the decision theoretic point of view – a cautionary note on updating imprecise priors. In: [6], pages 31-45.

[3] T. Augustin. Optimal decisions under complex uncertainty – basic notions and a general algorithm for data-based decision making with partial prior knowledge described by interval probability. *ZAMM Zeitschrift für Angewandte Mathematik und Mechanik*, 10-11:1-10, 2004.

[4] T. Augustin. Generalized basic probability assignments. Conditionally accepted for the *International Journal of General Systems*, 2005.

[5] J.O. Berger. Statistical Decision Theory and Bayesian Analysis. (2nd edition). Springer. New York, 1984.

[6] J.M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *ISIPTA 03. Proceedings of the Third International Symposium on Imprecises Probabbilities and Their Applications*, Lugano, Switzerland, July 2003. Carleton Scientific.

[7] J.-M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39:123-150, 2005.

[8] S. Blumenthal. Multinomial sampling with partially categorized data. *Journal of the American Statistical Association*, 63:542-551, 1968.

[9] A. Chateauneuf, M. Cohen, and I. Meilijson. New tools to better model behavior under risk and uncertainty: An overview. *Finance*, 18:25-46, 1997.

[10] F.P.A. Coolen and T. Augustin. Learning from multinomial data: a nonparametric predictive alternative to the Imprecise Dirichlet Model. Conditionally accepted for [14].

[11] G. de Cooman, F.G. Cozman, and S. Moral, and P. Walley, editors, *ISIPTA 99. Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, Zwijnaarde, Belgium, June 1999.

[12] G. de Cooman, T.L. Fine, and T. Seidenfeld, editors, *ISIPTA 01. Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, Ithaca, USA, June 2001. Shaker Publishing.

[13] G. de Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159:75125, 2004.

[14] F. G. Cozman, R. Nau, and T. Seidenfeld, editors, *ISIPTA 05: Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications*, Pittsburgh (Carnegie Mellon), USA, July 2005. Brightdocs.

[15] M. DeGroot. *Optimal Statistical Decisions.* McGraw-Hill, New York, 1970.

[16] A.P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38:325-339, 1967.

[17] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75:643-669, 1961.

[18] P. Gärdenfors and N.-E. Sahlin. Unreliable probabilities, risk taking, and decision making. *Synthese*, 53:361–386, 1982.

[19] P. Ghirardato and M. Marinacci. Risk, ambiguity, and the separation of utility and beliefs. *Mathematics of Operations Research*, 26:864-890, 2001.

[20] P. H. Giang and P. P. Shenoy. Statistical Decisions Using Likelihood Information Without Prior Probabilities. *Darwiche and N. Friedman, editors, Uncertainty in Artificial Intelligence, 2002*, pages 170-178, San Francisco, CA, 2002. Morgan Kaufmann.

[21] P. H. Giang and P. P. Shenoy. Decision making with partially consonant belief functions. *U. Kjærulff and C. Meek, editors, Uncertainty in Artificial Intelligence, 2003*, pages 272-280, San Francisco, CA, 2003. Morgan Kaufmann.

[22] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

[23] P. D. Grünwald and J. Halpern. When ignorance is bliss. In M. Chickering and J. Halpern, editors, *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence UAI 2004* , Banff, Canada, 2004, AUAI Press.

[24] D. F. Heitjan and D. B. Rubin. Ignorability and Coarse Data *The Annals of Statistics, Vol. 19*, 4:2244-2253, 1991.

[25] D. F. Heitjan. Ignorability and Coarse Data: Some Biomedical Examples. *Biometrics, Vol. 49*, 4:1099-1109, 1993.

[26] J. L. Hodges and E. Lehmann. The use of previous experience in reaching statistical decisions. *The Annals of Mathematical Statistics*, 23:396-407, 1952.

[27] J. Y. Jaffray. Rational decision making with imprecise probabilities. In [11], pages 324-332.

[28] E. Kofler and G. Menges. Entscheidungen bei unvollständiger Information. *Lecture Notes in Economics and Mathematical Systems, 136*, Springer, Berlin, 1976.

[29] E. Kofler. *Prognosen und Stabilität bei unvollständiger Information.* Campus, Frankfurt/Main, 1989.

[30] V. P. Kuznetsov. *Interval Statistical Models.* Radio and Communication, Moscow, 1991 (in Russian).

[31] I. Levi. On indeterminate probabilities. *Journal of Philosophy*, 71:391-418, 1974.

[32] S. Maier. *Entscheidung und Analyse bei unvollständiger Information.* Diploma thesis, Universitity of Munich, 2004.

[33] R. F. Nau. Uncertainty aversion with second-order probabilities and utilities. In [12], pages 273-283.

[34] H. T. Nguyen and E. A. Walker. On decision making using belief functions. In R. Y. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 311-330. Wiley, New York, 1994.

[35] M. Schervish, T. Seidenfeld, J. Kadane, J., and I. Levi. Extensions of expected utility theory and some limitations of pairwise comparisons. In [6], pages 496-510.

[36] J. Schubert. On ρ in a decision-theoretic apparatus of Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 13:185-200, 1995.

[37] T. Seidenfeld. A contrast between two decision rules for use with (convex) sets of probabilities: - Maximin Versus E-Admissibilty. *Synthese*, 140:69-88, 2004.

[38] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[39] T. M. Strat. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, 4:391-418, 1990.

[40] F. Tonon, A. Bernardini, and A. Mammino. Determination of parameter range in rock engineering by means of random set theory. *Reliability Engeneering and System Safety*, 70:241-261, 2000.

[41] M. C. M. Troffaes. Decisions making with imprecise probabilities: a short review. *F.G. Cozman editor, The SIPTA Newsletter, Vol. 2 No. 1: Society for Imprecise Probability Theory and Applications,* Manno, Switzerland. `http://www.sipta.org,` pages 4-7, Dec. 2004.

[42] L. V. Utkin. Extensions of belief functions and possibility distributions by using the imprecise Dirichlet model. *Fuzzy Sets and Systems*, 2005. In press.

[43] L. V. Utkin and T. Augustin. Decision making with imprecise second-order probabilities. In [6], pages 545-559.

[44] L. V. Utkin and T. Augustin (2005): Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. Conditionally accepted for [14].

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

[46] P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3-57, 1996 (with discussion).

[47] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung, volume I: Intervallwahrscheinlichkeit als umfassendes Konzept.* Physika, Heidelberg, 2001.

[48] K. Weichselberger. The logical concept of probability and statistical inference. Conditionally accepted for [14].

[49] R. R. Yager. Decision making under Dempster-Shafer uncertainties. *International Journal of General Systems*, 20:233-245, 1992.

[50] M. Zaffalon. Exact credal treatment of missing data. *Journal of Statistical Planning and Inference*, 105:105-122, 2002.