



Decision making under incomplete data using the imprecise Dirichlet model

L.V. Utkin ^{a,*}, Th. Augustin ^b

^a *Department of Computer Science, Forest Technical Academy,
Institutski per. 5, 194021 St. Petersburg, Russian Federation*

^b *Department of Statistics, University of Munich, Germany*

Received 15 December 2005; received in revised form 30 June 2006; accepted 31 July 2006

Available online 25 September 2006

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, including missing data. We derive efficient algorithms and discuss properties of the optimal solutions with respect to several criteria, including Gamma-maximinity and *E*-admissibility. In the case of precise data and pure actions the former surprisingly leads us 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.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Belief functions; Coarse data; Decision making; *E*-admissibility; Imprecise Dirichlet model (IDM); Imprecise probabilities; Incomplete data; Interval probability; Interval statistical models; Missing or set-valued statistical data

* Corresponding author.

E-mail addresses: lvu@utkin.usr.etu.spb.ru (L.V. Utkin), thomas@stat.uni-muenchen.de (Th. Augustin).

1. Introduction

When applying the theoretical framework of decision theory to substantial science problems, decision makers typically have only limited information about probability distributions involved in the problem, and so the analysis is associated with large uncertainty. As a result, decision makers are frequently confronted with the problem that the very demanding and strong conditions of the classical probability calculus, and the decision models based on it, are not satisfied. Then it is indispensable 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 sophisticated uncertainty representation generalizing the common ('classical') concept of probability can be found in the literature, including Dempster–Shafer structures [18,42], interval-valued probabilities [51], imprecise probabilities [33,49], see also [6,12,13]. The corresponding decision making models have been developed in accordance with the different types of the uncertainty representation (e.g. [1,3,20,36,23,37,43,47,53]). In contrast to standard decision theory, these models allow to handle *partial* information about the stochastic behavior of the states of nature.

Here we explicitly take 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 also seems to be suitable for processing expert judgements, as long as they are based on independent sources of information.

A particularly attractive feature of our method is that it will prove to be able to incorporate even set-valued observations, i.e., to handle situations where the corresponding category cannot 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. [50]; 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 cannot 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 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

Download English Version:

<https://daneshyari.com/en/article/397566>

Download Persian Version:

<https://daneshyari.com/article/397566>

[Daneshyari.com](https://daneshyari.com)