The Apparent Arbitrariness of Second-Order Probability Distributions

The Apparent Arbitrariness of Second-Order Probability Distributions

David Sundgren

UNIVERSITET STOCKHOLMS

Stockholm University

© David Sundgren, Stockholm 2011

ISSN 1101-8526
ISBN 978-91-7447-184-7

Printed in Sweden by Universitetsservice US-AB, Stockholm 2011

Distributor: Department of Computer and Systems Sciences, Stockholm University

Δεσμῶν δὲ κὰλλιστος ὃς ἂν αὔτὸν καὶ τὰ συνδούμενα ὅτι μάλιστα ἓν ποιῆ

Πλάτων, «Τίμαιος»

Abstract

Adequate representation of imprecise probabilities is a crucial and non-trivial problem in decision analysis. Second-order probability distributions is the model for imprecise probabilities whose merits are discussed in this thesis.

That imprecise probabilities may be represented by second-order probability distributions is well known but there has been little attention to specific distributions. Since different probability distributions have different properties, the study of the desired properties of models of imprecise probabilities with respect to second-order models require analysis of particular second-order distributions.

An often held objection to second-order probabilities is the apparent arbitrariness in the choice of distribution. We find some evidence that the structure of second-order distributions is an important factor that prohibits arbitrary choice of distributions. In particular, the properties of two second-order distributions are investigated; the uniform joint distribution and a variant of the Dirichlet distribution that has the property of being the normalised product of its own marginal distributions.

The joint uniform distribution is in this thesis shown to have marginal distributions that belie the supposed non-informativeness of a uniform distribution. On the other hand, the modified Dirichlet distribution discovered here has its information content evenly divided among the joint and marginal distributions in that the total correlation of the variables is minimal.

It is also argued in the thesis that discrete distributions, as opposed to the continuous distributions mentioned above, would have the advantage of providing a natural setting for updating of lower bounds, and computation of expected utility is made more efficient.

Summarium

In placitorum scrutatione maxima et mehercle minime levis difficultas eo spectat, quomodo probabilitates dubiae bene ostendantur. In hac thesi de utilitate distributionum probabilitatum secundi ordinis disseremus, in quantum ad probabilitates dubias ostendendas valeant.

Omnibus fere notum est probabilitates dubias ostendi posse per distributiones probabilitatum secundi ordinis, sed pauci operam distributionibus singulis operam contulerunt. Cum tamen distributiones probabilitatum valde inter se diversae sint, si quis proprietatibus desideratis probabilitatum dubiarum secundi ordinis studium conferre vult, primum debet quasdam praescriptas distributiones secundi ordinis investigare.

Sed fortasse, quod saepenumero fieri solet, quispiam dixerit probabilitates secundi ordinis nulla, ut videtur, ratione habita quasi vagari quoad delectum distributionis. Nos tamen nonnulla indicia comperimus quibus freti confirmare audemus ipsam formam distributionum secundi ordinis multum valere ad praedictum distributionum secundi ordinis delectum rationabiliter peragendum. Imprimis proprietates duarum distributionum secundi ordinis investigabimus, nimirum distributionis uniformis coniunctae et alterius cuiusdam speciei distributionis quae ‘Dirichleti’ vocatur, quae ex ipsius distributionibus marginalibus ad normam correcta oritur.

In hac thesi probamus illam coniunctam uniformem distributionem continere distributiones marginales eius modi quae illos refellant qui negant distributionem uniformem quicquam alicuius momenti afferre. Attamen in illa distributione Dirichleti paulo mutata, quam hoc loco patefacimus, omnia aequaliter inter coniunctas et marginales distributiones divisa sunt, in quantum tota ratio quae inter variantia intercessit ad minimum reducitur.

Insuper in hac thesi confirmamus distributiones discretas potius quam antedictas distributiones continuas in hoc utiliores esse, quod per eas limites inferiores in melius mutare licet, et beneficia exspectata accuratius computari possunt.

Sammanfattning

Adekvat representation av osäkra eller imprecisa sannolikheter är ett avgörande och icke-trivialt problem i beslutsanalys. I denna avhandling diskuteras förtjänsterna hos andra ordningens sannolikheter som en modell för imprecisa sannolikheter.

Att imprecisa sannolikheter kan representeras med andra ordningens sannolikheter är välkänt, men hittills har särskilda andra ordningens fördelningar inte ägnats någon större uppmärksamhet. Då olika sannolikhetsfördelningar har olika egenskaper kräver studiet av önskvärda egenskaper hos modeller för imprecisa sannolikheter en granskning av specifika andra ordningens fördelningar.

Den godtycklighet som tycks vidhäfta valet av andra ordningens fördelning är en ofta förekommande invändning mot andra ordningens sannolikhetsfördelningar. Vi finner vissa belägg för att strukturen hos andra ordningens fördelningar är en omständighet som hindrar godtyckligt val av fördelningar. I synnerhet undersöks egenskaper hos två andra ordningens fördelningar; den likformiga simultana fördelningen och en variant av Dirichletfördelningen med egenskapen att vara lika med den normaliserade produkten av sina egna marginalfördelningar.

Den likformiga simultana fördelningen visas i avhandlingen ha marginalfördelningar som motsäger den förmodat icke-informativa strukturen hos en likformig fördelning. Å andra sidan gäller för den modifierade Dirchletfördelningen som upptäckts här att informationsinnehållet är jämnt fördelat mellan den simultana fördelningen och marginalfördelningarna; den totala korrelationen mellan variablerna är minimal.

Det hävdas också i avhandlingen att diskreta sannolikhetsfördelningar i motsats till de kontinuerliga fördelningar som nämnts ovan har fördelen att utgöra en naturlig miljö för uppdatering av undre gränser och dessutom tillåta en mer effektiv beräkning av förväntad nytta.

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I	M. Danielson, L. Ekenberg, K. Hansson, J. Idefeldt, A. Larsson, M. Påhlman, A. Riabacke and D. Sundgren. Cross-disciplinary research in analytic decision support systems. In *ITI 2006: proceedings of the 28th International Conference on Information Technology Interfaces, Cavtat, Croatia. Zagreb: University Computing Centre SRCE, University of Zagreb; 2006. p. 123-128.*, 2006.
II	M. Danielson, L. Ekenberg and D. Sundgren. Structure information in decision trees and similar formalisms. In *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference. Menlo Park, CA: AAAI Press; 2007. p. 62-67*, 2007.
III	D. Sundgren, M. Danielson, L. Ekenberg. Warp effects on calculating interval probabilities. In *International Journal of Approximate Reasoning. 2009;50(9):1360-1368.* 2009.
IV	D. Sundgren, L. Ekenberg, M. Danielson. Some properties of aggregated distributions over expected values. In *MICAI 2008: Advances in Artificial Intelligence: 7th Mexican International Conference on Artificial Intelligence, Atizapán de Zaragoza, Mexico, October 27-31, 2008, Proceedings. Berlin, Heidelberg: Springer; 2008. p. 699-709. Lecture Notes in Computer Science, 5317.* 2008.
V	Sundgren D, Ekenberg L, Danielson M. Shifted Dirichlet Distributions as Second-Order Probability Distributions that Factors into Marginals. In: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications. 2009. p. 405-410.
VI	Sundgren D. Expected Utility from Multinomial Second-order Probability Distributions. Mexico City: Centro de Innovacíon y Desarrollo Tecnológico en Cómputo, Instituto Politécnico Nacional; Polibits Journal of Research and Development in Computer Science and Enigineering. 2010;(42):71-75.

Reprints were made with permission from the publishers.

Innehåll

1. Background

1.1 Decision Analysis

In this thesis, second-order probability as a model for imprecise probability is considered in the context of decision analysis. Be it due to limited resources, conflicting evidence or the uncertainty inherent in complex systems, that imprecise probabilities are needed in decision analysis is hardly controversial [25, 55, 56, 16]. But the choice of model for imprecise probability is a source of controversy and what follows the particular theory called second-order probability is investigated and compared to some of the alternatives.

The need for adequate representation of imprecise probability in decision support systems is argued for in paper I of the present thesis. The authors point out that it is when confronting complicated decision problems that decision support systems are needed in the first place since no one needs much help solving a simple problem. Then again, what help is such as system if it offers no assistance with the difficult parts of the problem? And it is as a rule the analysis of the decision problem that poses difficulties. The topic of this thesis is limited to the problem of handling probabilities.

Although the thesis is to a large extent theoretical in nature, the motivation and aim is practical. In decision problems that are so complicated that probability is needed in the first place there is also need for imprecise probability. There is a plethora of mathematical representations of imprecise probability that suffer from weaknesses that impair their practical applicability. In this work one of these models of imprecise probability is investigated with the aim of shedding light on some of its perceived weaknesses and to demonstrate that this model may be more usable and efficient than previously believed. The results of the present thesis are then significant steps towards the indeed practical and pressing problem of high quality decision making.

It is usually hard for humans to think in terms of probability, see e.g. [47]. For this reason it is important that decision support systems give adequate guidance for the user when it comes to assessing probability values, precise or imprecise. Since the probabilities of the possible outcomes of an event are in general entangled with each other through dependencies and it is hard to asses the situation in its entirety from a bird's eye view it would be overly optimistic to ask the user of a decision support system to enter his or her probability estimates and expect a high quality end result unless the user is an expert in probability theory.

Due to lack of sufficient amounts of data, in most real-life decision situations, it is impossible for the decision maker to aggregate precise numerical probability values. And there appears to be a psychological tendency to avoid precise probability estimates. This reluctance could indeed issue from a sound awareness that such precise values do not strictly speaking reflect the truth, or at least an admittance that the hesitant individual does not understand the nature of probability enough to make statements of precise probability. There is also the possibility that due the chaotic nature of complex systems there are problems where uncertainty is inherent to the degree of making all precise probability statements false, where an exact probability value simply does not exist. In either case decision support systems need to allow the user to enter imprecise probability statements.

Care must be taken, though, in the choice of model for imprecise probability. An imprecise probability value could be viewed as a collection of all probability statements that are possible given some limited information, but when implementing that idea in a specific theory it is tempting to ignore what is actually possible given the probabilistic structure of the problem. By taking into account the structure of the decision problem and the information inherent therein the input process required of the user can be simplified and the quality of the output enhanced. As for input, structural considerations may help the decision maker break down the problem into parts that are more easily comprehended. And the output is dependent on discrimination of decision alternatives, a discrimination that can be made clearer by using structural information. The less of probability theory that is incorporated in the system, the more is demanded of the user. Probability is a complicated concept that can confuse the best of us.

Thus there ought be a component in decision support systems that aids the user in structuring the problem with regards to structural properties such as dependencies and in formulating consistent probability values. Such a component would require a model for imprecise probability that has the capacity to take into account factors such as dependency. In probability theory there is already the machinery for dealing with dependency, an imprecise probability theory based on probability theory itself would inherit the tools necessary for modeling the complex situations that are encountered in decision problems.

What is meant by basing imprecise probability on probability theory is that when there is uncertainty as to the precise value of a probability one can assign probability values to the possible probability values, "the probability that the probability of this loaded die showing six is 0.2 is 0.7 but the probability that the probability that the die shows six is 0.1 is 0.3". Since probability values are assigned to probability values, the term *second-order probability* is employed.

Second-order probability has hitherto seen little practical use in decision analysis largely since a set of beliefs concerning imprecise probabilities on the face of it is consistent with a large, even infinite, number of second-order probability distributions. In this thesis it is shown that the choice of second-order distribution has consequences that a decision maker ignores at his own peril.

The reasoning here does not rely on any particular philosophical or psychological theory on the nature of probability though it is suggested that some models are better suited to some views on probability than others. However, a fundamental assumption of this work is the decision theoretical principle of maximising expected utility, introduced in [53]. In its original formulation probabilities are considered as objective but subjective expected utility was treated by [45], (first edition published in 1954).

For practical use, the principle of maximising expected utility is a decision rule. Assume that there are a number of decision alternatives, one of which is to be selected as optimal. The outcome of a decision alternative depends on the outcome of random events, say that there are n different relevant mutually exclusive outcomes to a certain decision alternative. Outcome i occurs with probability p_i and has utility u_i. The expected utility of the decision alternative in question is then $\sum_{i=1}^{n} p_i u_i$ and the alternative with the highest maximal expected utility is chosen.

Either objective or subjective, the principle of maximising expected utility demands precise values for probabilities p_i and utilities u_i. But in situations that are so complex that decision analysis is needed in the first place it is rare that such precise values can be given, hence the need for imprecise probabilities and utilities. For imprecise utilities, not further covered in this thesis, see e.g. [28, 2, 3]. With imprecise probabilities the principle of maximising expected utility cannot be employed without modification of some sort.

Models of imprecise probability may with some simplification be divided into two classes, interval-based models and hierarchical models. In an interval-based theory of imprecise probability such as those of [32, 33, 48, 55, 59, 58], the probability of an event is represented by two numbers, the lowest and highest possible value of the probability. The differences between such models lie largely in philosophical assumptions on the nature of probability and psychological theories on how the extreme probability values are to be interpreted, and whether the interval extreme values are to be denoted probabilities or some other related concept. In hierarchical models such as described in e.g. [9, 39, 21, 41, 61, 60, 19, 18, 51, 49, 43], there are also the extreme probability values as in non-hierarchical models, but each probability value in the interval is weighed. The hierarchy is in that the weighing is on an order higher than the probabilities of the events. The various hierarchical models are distinguished mainly by the constitution of the weights. The model argued for in this thesis, second-order probability, is a hierarchical model where the weights are interpreted as probabilities.

A system for decision analysis can be no more efficient than its decision rule, the decision rule is indispensable and decisive. Even more important than efficiency would be the degree to which a decision rule utilises available and relevant information and is adequately sensitive to changes in the decision maker's beliefs. Different models of imprecise probability naturally entail different decision rules although all rules described here are variants of maximising expected utility. When imprecise probabilities are modelled in a non-hierarchical fashion, such as lower probabilities [48] or lower previsions [55], expected utility takes the form of an interval. In the case that the lowest expected utility of one decision alternative is higher then the highest expected utility of all the other alternatives, the first mentioned alternative is said to dominate the others and is clearly optimal. Usually, though, expected utility intervals overlap and then two mutually complementary decision rules come into play, maximin, intended for the pessimist, where the maximum of the minimal expected utilities is sought, or maximax for the optimistically inclined, where the decision alternative that has the highest maximum value of expected utility is chosen, see e.g. [21]. An alternative approach is to contract the intervals of probabilities and utilities and with them the intervals of expected utility until domination occurs as in [11].

As for hierarchical models, a version of the maximin rule is in [9] suggested for second-order possibility distributions, $\underline{P}$-maximin, in [18] the concept of generalised expected utility, using the mean values of probabilities and utilities as the precise values required for computation of expected utility, is introduced as a decision rule based on second-order probabilities. The hierarchical nature of these model for probabilities and utilities is not fully reflected in the respective interpretations of expected utility. Specifically, in the hierarchical models mentioned above there is second-order information that could be used to discriminate between values of expected utility, but only the smallest and largest possible values of expected utility are accounted for and that without consideration of the weights of these values.

With given second-order probabilities and utilities a distribution expected utility is induced, and neither in pure interval-based models nor in hierarchical models the stochastic nature of expected utility has been exploited until recently. In contrast, distribution of expected utility could serve as a basis for a second-order decision rule. for instance, in [6] the difference $E(A) - E(B)$ between distributions of expected utility is suggested for distinguishing between alternatives A and B.

Another possibility for pairwise comparison of decision alternatives with discrete second-order distributions is to compute the probability that alternative A yields a higher expected utility than alternative B, see paper VI. Let z and x be possible values of expected utility for alternatives A and B, and $h_A(x)$ and $h_B(x)$ be the probability density functions for expected utility. Since the probability of the conjunction of independent events is the product of their respective probabilities and the probability of the disjunction of mutually exclusive events is the sum of the probabilities, if the expected utility of the two alternatives are independent, the probability could be computed as $\sum_z \sum_{x<z} h_A(z) h_B(x)$. Note that such a probability value does not constitute a decision rule in itself, a limit would have to be set so that an alternative is ad-

vocated if the probability of higher expected utility is at least e.g 80 %. This limit would reasonably differ with different types of decision. For continuous distributions, integrals would be used instead of sums, but the values would have to computed numerically since the integrals would be nearly impossible to evaluate exactly.

1.2 Imprecise Probability

The concept of imprecise probability is no more specified than to contain probabilities that are not precise, hence it can be realised in different ways. This thesis is primarily concerned with second-order probability, which is one possible instantiation of imprecise probability. Before describing second-order probabilities in a little more detail, let us give a background to imprecise probabilities by mentioning some alternative theories. Many theories of imprecise probability are based on sets of probability measures, also called credal sets. A credal set is informally a set of probability distributions. Such a set is usually restricted by lower (or, equivalently, upper) bounds of probabilities, and the demand that the set is convex. The intuition appears to be that instead of choosing precise probabilities one uses the set of all probability distributions that are consistent with the beliefs of an agent. The considerable size of the credal set makes it rather unwieldy, but according to [52] most of the Dempster-Shafer theory [46] holds in a subset in the form of lattice. Theories of this kind include Choquet capacities [8], lower probabilities [48] and lower previsions [55]. These theories are accessibly summed up in [10].

This thesis is devoted to the study of a particular form of imprecise probability, *second-order probability*. Second-order probability is an example of a hierarchical model in that there are two levels of probability. The first-order level is the probability of an event that is relevant for the decision at hand and the second-order level is the probability that the before mentioned probability has a specific value. The introduction of second-order probability naturally raises the question of whether third- and higher order probabilities may be useful or even necessary.

The issue of higher-order probability is dealt with below on page 32 in Section 1.3.6 but for the sake of clarification it is in order to shed some light on the matter at this point. Simply put, second-order probability is needed for modelling of uncertain probability, and third-order probability would model the uncertainty of the uncertainty. While there definitely is room for questioning the degree of certainty as regards to uncertainty, firstly this higher-order uncertainty would normally be less then the original uncertainty, secondly the higher up in this hierarchy a person goes the harder it becomes to understand what is happening as the air gets thinner, as it were.

While the first argument is that higher-order probability makes little difference in practice the second argument claims that higher-order probabilities from some order and up are beyond the capacity of human subjects. To summarise with an example, I believe that the probability of an event is between 0.2 and 0.5, with some second-order distribution for the probability of values in between. Questioned, I might admit that the lower bound might be as low as 0.15 and the higher bound might in fact be as much as 0.6, although it would be unlikely (otherwise I would revise my original bounds). My third-order distribution would then differentiate between the second-order distributions that are consistent with these values. But since the difference in values is relatively small and and the extremes unlikely, the end result when it comes to modelling my beliefs would not be very different from that modelled by the original second-order distribution. And when it comes to fourth-order reasoning, when questioned on how certain I am about my third-order probabilities, it should not be held against me if I threw up my hands in the air and gave up. Admittedly, the above reasoning concerning higher-order probability is vague and intuitive. The facts would have to be determined by further research.

Hierarchical models in general do not necessarily involve probability, they can rely on other measures at either the first or second level. Indeed, the hierarchical concept need not in theory be limited to two levels, although deeper levels of hierarchy is rarely found in the literature. Below we describe the outlines of some hierarchical models in order to give a backdrop to second-order probability, and to invite comparisons. These are alternative theories of imprecise probability and they do not build on each other. Hence full understanding of the different often advanced theories is not required for the comprehension of second-order probability.

Before the rather detailed exposition of hierarchical models, let us mention some milestones not further mentioned in the thesis. Capacities of order 2, [8], marks a beginning in 1953-1954, although capacities of order 2 was not used for modelling imprecise probabilities until 1994, [15]. With [48] and [24], in the early nineteen-sixties, classes of probability measures began to be used for interval-valued probability functions. The Dempster-Shafer theory was incepted in 1967 with [14] and was completed in 1976 by [46].

1.3 Hierarchical Models

The notion that some members of a credal set is more likely, reasonable, reliable or probable than others give rise to hierarchical models, sometimes also called second-order or second-order uncertainty models. Such models have in common that they allow for discrimination of probability distributions, different distributions are assigned different weights. The possibilistic hierarchy of [9], epistemic reliability [21], likelihood discrimination [41] and fuzzy probabilities [61, 60] are hierarchical models that are not restricted to second-order probability distributions.

1.3.1 Possibilistic hierarchy

De Cooman and Walley [9] present a *possibilistic* hierarchical model for behaviour under uncertainty. The *possibility* space Ω is the set of possible states of the world, assumed to be mutually exclusive and exhaustive. A *gamble* X is a bounded, real-valued function on Ω. The reward $X(\omega)$ is assumed to be on a linear utility scale.

The goal is to model uncertainty about probabilities; let the *subject* be a person that has made some subjective assessments about probabilities of the states in Ω. The *modeller* wishes to express his uncertainty about the subjects probabilities.

Let $R(X,x)$ denote the event that the subject will refuse to buy the gamble X for the price x. Then the modellers second-order uncertainty may be modelled by his propensity to bet against the event $R(X,x)$. For any real x, the *buying function* $\beta_X(x)$ is defined as the modellers *infimum acceptable betting rate* for betting against $R(X,x)$. Equivalently, $\beta_X(x)$ is the modellers *upper probability* of $R(X,x)$.

Example 1 *[9] If the modeller only knows that the subjects lower prevision for X is c (i.e. the subject is committed to buy X for any price smaller than c) the corresponding buying function is* $\beta_X(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c\,. \end{cases}$

Analogously, a *selling function* σ_X is the modellers upper probability of the event $R'(X,x)$ that the subject will refuse to sell X for x units.

The authors of [9] require that the buying functions satisfy a set B of axioms, one of which is that $\beta_{X+Y}(x+y) \leq \max\{\beta_X(x), \beta_Y(y)\}$ for all real numbers x, y, c and all gambles $X, Y \in \mathscr{K}$. It is this axiom that renders the theory *possibilistic*. A *possibility measure* $\overline{P}$ is characterised by the property that for any two events A and B, $\overline{P}(A \cup B) = \max\{\overline{P}(A), \overline{P}(B)\}$.

Buying functions that satisfy the axiom set B are called *possibilistically coherent*.

Since buying and selling functions are conjugate, $\sigma_X(x) = \beta_{-X}(-x)$, it is sufficient to study the theory of either buying or selling functions. The information in buying and selling functions may be incorporated in *price functions*.

Definition 1 *[9] Given possibilistically coherent buying function β_X defined for all X in a linear space of gambles $\mathscr{K}$ the* price function ρ_X *is defined by*

$$\rho_X(x) = \min\{\beta_X(x), \beta_{-X}(-x)\} = \min\{\beta_X(x), \sigma_X(x)\} \tag{1.1}$$

for every X in $\mathscr{K}$ and all real numbers x.

Intuitively, the price function serves to combine the information from buying and selling functions into a measure of imprecise probability. The value $\rho_X(x)$ could be seen as determining to which degree it might be that the probability of X occurring is in fact x in the same way as a fuzzy membership function. The in a sense likeliest probability value has $\rho = 1$ and for impossible probability values x $\rho(x) = 0$. In [9] two examples from [57] are shown , one of which is that the subject states that "the probability of A is about 20%"'. From corresponding betting rates a price function ρ is formed such that $\rho_A(x) = 1 - 10|x - 0.2|$ if $0.1 < x < 0.3$ and $\rho_A(x) = 0$ for all other values of x.

Buying and selling functions can be recovered from price functions through a *mode* of the price function, i.e. m_X such that $\rho_X(m_X) = 1$.

$$\beta_X(x) = \begin{cases} \rho_X(x), & \text{if } x < m_X \\ 1, & \text{if } x \geq m_X \,. \end{cases} \tag{1.2}$$

$$\sigma_X(x) = \begin{cases} \rho_X(x), & \text{if } x > m_X \\ 1, & \text{if } x \leq m_X \end{cases}$$

If the subject has a *fair price* $P(X)$, such that $P(X)$ is both his supremum acceptable buying price and his infimum acceptable buying price, then $\rho_X(x)$ may be interpreted as the modellers upper probability that $P(X) = x$.

Example 2 *[9] Suppose that the subject reveals both his lower prevision c and his upper prevision $d, (d \geq c)$ for X. Then this information is modelled by the price function*

$$\rho_X(x) = \begin{cases} 0, & \text{if } x < c \text{ or } x > d \\ 1, & \text{if } c \leq x \leq d \,. \end{cases}$$

There is in [9] a corresponding set of axioms P derived from B, and these axiom sets are in [9] proved to be equivalent.

As with buying functions, price functions are called *possibilistically coherent* if they satisfy the axioms.

Second-order possibility distributions

Price functions ρ_X concern just one gamble at a time; they are *local models*. Possibility distributions are designed to model the subjects overall probability model. Possibilistically coherent price functions are exactly the price functions that can be generated by this global model.

Assume that the subject is Bayesian, whose actions are determined by a *linear prevision*. A linear prevision P is a real-valued function such that $P(X)$ is defined for all gambles on the possibility space Ω, which satisfies the convexity axiom $P(X) \geq \inf[X]$ and the linearity axiom $P(X+Y) = P(X) + P(Y)$. The value $P(X)$ is interpreted as the subject's fair price for the gamble X. Every linear prevision is the expectation operator of some probability measure. Let $\mathscr{P}$ denote the set of all linear previsions on Ω.

In [9] $\pi : \mathscr{P} \to [0,1]$ such that $\sup\{\pi(P) : P \in \mathscr{P}\} = \infty$ is defined to be a *second-order possibility distribution*. The number $\pi(P)$ is interpreted as the upper probability for the modeller that the subject's true linear prevision P_T is P. And $\pi(\mathscr{R})$ measures how plausible it is that $P_T \in \mathscr{R}$.

A problem with the generous definition of second-order possibility distributions appears to be that it does not preclude that $\pi(x) < \pi(y)$ when $y < x$. Even granted that a possibility distributions is not a probability distribution, the possibility of $X \leq x$ should be larger than the possibility of $X \leq y < x$.

A price function can be generated from a possibility distribution by

$$\rho_X(x) = \pi(\{P \in \mathscr{P} : P(X) = x\}) = \sup\{\pi(P) : P \in \mathscr{P}, P(X) = x\}, \quad (1.3)$$

with $\rho_X(x) = 0$ if the set in (1.3) is empty.

Similarly, π generates a buying function $\beta_X(x) = \pi(\{P \in \mathscr{P} : P(X) \leq x\})$ and a selling function $\sigma_X(x) = \pi(\{P \in \mathscr{P} : P(X) \geq x\})$.

So an alternative interpretation of $\beta_X(x)$ is the upper probability that the subject's fair price $P_T(X)$ satisfies $P_T(X) \leq x$.

It is pointed out in [9] that a limitation of second-order possibility distributions is the the subject is assumed to be Bayesian. However, price functions does not rely on the subject being Bayesian, and just as possibility distributions generate price functions, price functions generate possibility distributions. But $\pi(P)$ can, according to [9], be interpreted as the modeller's infimum upper probability, over all games X in $\mathscr{K}$, that the subject will refuse to buy X for the price $P(x)$, an interpretation that does not rely on the assumption that the subject is Bayesian.

Decision making with second-order possibility distributions

The authors of [9] suppose that the modeller is to choose a decision alternative, or action, from a set $\mathscr{A}$ and that the consequences of an action $\mathfrak{a} \in \mathscr{A}$ can be evaluated by a bounded linear utility function $X_{\mathfrak{a}}(\omega)$, where ω is the true state, that measure's the modeller's expected utility from choosing $\mathfrak{a}$.

To choose a unique maximal action the authors of [9] suggests maximising $\underline{P}(X_{\mathfrak{a}})$, yielding a so called $\underline{P}$*-maximin* action. If $\mathscr{K}$ is compact and convex, there is at least one $\underline{P}$-maximin action. When $\underline{P}$ is vacuous the definition of $\underline{P}$-maximin action agrees with the standard notion of maximin. And if $\underline{P}$ is defined on an algebra of events and is a two-monotone Choquet capacity, and is further extended to gambles X by defining $\underline{P}(X)$ as the Choquet integral of X with respect to the capacity, then an actions is $\underline{P}$-maximin if and only if it maximises Choquet expected utility.

However, referring to Section 5.6 of [55], the authors point out that they do not suggest that either using $\underline{P}$-maximin or maximising Choquet expected utility is a good way of making decisions. There is in [9] no further motivation for dismissing $\underline{P}$-maximin and Choquet expected utility for singling out optimal actions.

Another way of making decisions in the possibilistic hierarchical model is suggested in [9] through the fact that the price function ρ_A of an event A can be seen as a *fuzzy probability*. In fact, there is not one, but several different ways of ranking actions based on fuzzy expected utility. The authors of [9] find it, without further explanation, reasonable, and in line with [44] to use the price function ρ_X as fuzzy expected utility, rather than the ε_X of [57].

One of the ranking methods mentioned in [9] is the centre-of-mass method [44]; maximise the centre of mass of the fuzzy expected utility ε_X $\int x\varepsilon_X(x)\,dx / \int \varepsilon_X(x)\,dx$. There is a striking resemblance with the centroid method in [18]. In the centre-of-mass method a fuzzy number is integrated, but in the centroid method the second-order distributions of utilities and probabilities are integrated separately and the expected utility is computed as the inner product of the centroids. The centroid of a utility or a probability vector u or p in the sense of [18] could as well be considered masses of centre since $\int u(x)\,dx = \int p(x)\,dx = 1$.

It is not clear how utility and possibility functions of a complex decision problem are combined to produce a price function of an action. We suggest using separate price functions for utilities and probabilities.

As remarked in [9], a second-order possibility distribution π may be interpreted as *epistemic reliability*, as in [21], described below in section 1.3.2, since $\pi(P)$ measures the reliability of the probability distribution P.

In conclusion, the possibilistic hierarchy is a solid framework for representation of uncertain probabilities that allow for e.g. interval-based probabilities as well as second-order models such as fuzzy probabilities. The possibilistic theory makes no claims about practical issues of implementation.

1.3.2 Epistemic reliability

In [21] Gärdenfors and Sahlin considers sets of probability measures $\mathscr{P}$ as a way to model second-order uncertainty by assigning each probability measure P in $\mathscr{P}$ a measure of *epistemic reliability*, $\rho(P)$.

Even though ρ is assumed to be real-valued and bounded both from below (when there is no second-order information about the elements of $\mathscr{P}$) and from above (representing complete information), it will only be necessary that the distributions in $\mathscr{P}$ can be ordered with respect to their epistemic reliability ρ.

Desired level of epistemic reliability

In the decision theory of [21], decision making consists of two steps; the first is to restrict P to a set $\mathscr{P}/\rho_0$ of probability measures with a minimum epistemic reliability, i.e. P with $\rho(P) \geq \rho_0$ for some appropriate choice of ρ_0, depending on the modeller's aversion to risk.

The next step is applying the maximin criterion for expected utilities; for each alternative a_i and each probability distribution in $\mathscr{P}/\rho_0$ the expected utility e_{ik} is computed in the ordinary way. The minimal expected utility of each alternative a_i relative to $\mathscr{P}/\rho_0$ is computed, and finally the alternative with the largest minimal expected utility is chosen. Gärdenfors and Sahlin compare their work to that of Levi [38], who requires that the set of permissible probability measures (corresponding to $\mathscr{P}/\rho_0$) are convex, and questions the need for convexity. A case could be made for convexity by looking at the work of [9].

If the probability distributions in $\mathscr{P}$ are probability measures as in section 1.3.1, the set $\mathscr{P}/\rho_0$ is equal to the *cut set* in [9]. Recall that full second-order possibility distributions represent possibilistically coherent price functions, and that the cut sets of a possibility distribution is required to be convex for the distribution to be full.

Again, the work of Gärdenfors and Sahlin in [21] is purely theoretical, paying little attention to practical details. But as opposed to the possibilistic hierarchy of [9], the concept of epistemic reliability in [21] is a pure second-order model. And in [21], the authors do include a method for ranking decision alternatives.

1.3.3 Likelihood discrimination

Another take on modelling second-order uncertainty is offered by Luce and Raiffa in [41]. Suppose that a decision maker has a set A of pure alternatives, giving a set of gambles depending on a Boolean algebra E of chance events. If a and b are gambles and α is an alternative in E, $a\alpha b$ denotes the gamble in which a is the outcome if α occurs and b if it does not. The set of all such gambles, including the alternatives in A, is denoted by G. If a and b are two gambles in G, $P(A,B)$ is the probability that an individual, corresponding to "the subject" in section 1.3.1, will prefer a to b.

The theory includes a probability $Q(\alpha,\beta)$, interpreted as the probability that α is more likely than β to occur. Hence Q is a second-order probability in the sense of being the probability of an event concerning probabilities. Q induces a relation on events;

$$Q(\alpha,\delta) \geq Q(\beta,\delta) \text{ and } Q(\delta,\alpha) \leq Q(\delta,\beta) \text{ for every } \delta \in E\,.$$

The context of the probability Q in [41] is the situation that the subject must decide between the two gambles $a\alpha b$ and $a\beta b$. Then there are two combinations of preference over a and b and the likelihoods of α and β that should lead to preference for $a\alpha b$ over $a\beta b$:

1. a is preferred to b, and α is deemed more likely to occur than β,
2. b is preferred to a, and β is deemed more likely to occur than α.

The probabilistic theory of utility in [41] is then characterised by a set of 11 axioms; the main result of [41] is that these axioms can not be fulfilled simultaneously.

In [41] it is assumed that there exists at least one real-valued *utility function* u and at least one real-valued function ϕ on E called the *subjective probability function*, which is, as opposed to Q, a first-order probability.

To be able to state the final results of impossibility in [41] without reference to independent events, the authors need axioms 9 and 10 to ensure that there is a dense set of independent events in E, though it is not clear how the axioms are used, since the results are not proven in the article.

It seems to be implied that decision making in this setting is to be made by maximising expected utility with utilities u and probabilities ϕ. Both u an dϕ can be calculated from the discriminatory probabilities P and Q, respectively. From the axioms it is established in [41] that the second-order discriminatory probability Q depends only on the difference of the subjective first-order probabilities of its two events.

$$Q(\alpha,\beta) = \begin{cases} 1/2+1/2[\phi(\alpha)-\phi(\beta)]^{\varepsilon}, & \text{if } \alpha \prec \beta \\ 1/2, & \text{if } \alpha \sim \beta \\ 1/2-1/2[\phi(\beta)-\phi(\alpha)]^{\varepsilon}, & \text{if } \beta \prec \alpha. \end{cases}, \quad (1.4)$$

where $\varepsilon > 0$ varies with the degree of discrimination.

When $0 < \varepsilon < 1$, ϕ can be expressed in terms of Q:

$$\phi(\alpha) = \begin{cases} 1/2+1/2[2Q(\alpha,\overline{\alpha})-1]^{1/\varepsilon}, & \text{if } Q(\alpha,\overline{\alpha}) > 1/2, \\ 1/2, & \text{if } Q(\alpha,\overline{\alpha}) = 1/2, \\ 1/2-1/2[1-2Q(\alpha,\overline{\alpha})]^{1/\varepsilon}, & \text{if } Q(\alpha,\overline{\alpha}) > 1/2 \end{cases} \quad (1.5)$$

A similar result holds for u and P;

$$P(a,b) = \begin{cases} 1/2+1/2[P(a^*,b^*)-P(b^*-a^*)][u(a)-u(b)]^{\varepsilon}, & \text{if } a \prec b, \\ 1/2, & \text{if } a \sim b, \\ 1/2-1/2[P(a^*,b^*)-P(b^*-a^*)][u(b)-u(a)]^{\varepsilon}, & \text{if } b \prec a \end{cases}, \quad (1.6)$$

and

$$u(a) = \begin{cases} \left[\frac{P(a,b^*)-P(b^*,a)}{P(a^*,b^*)-P(b^*,a^*)}\right]^{1/\varepsilon}, & \text{if } a \prec b^* \\ 1-\left[\frac{P(a^*,a)-P(a,a^*)}{P(a^*,b^*)-P(b^*,a^*)}\right]^{1/\varepsilon}, & \text{if } b^* \prec a \end{cases}, \quad (1.7)$$

where ε is the discrimination constant associated with Q and a^* and b^* are as in Axiom 5.

Impossibility

As their final conclusion, the authors of [41] states what they call an impossibility theorem; namely that the 11 axioms above cannot be satisfied simultaneously.

The argument is that when the subject can discriminate perfectly, there are at least three properties a,b and c which are perfectly discriminated with respect to preference, i.e. $P(a,b) = P(b,c) = P(a,c) = 1$. E.g. when a,b and c are certain rewards of \$10, \$5 and \$1, respectively. Perfect discrimination entails that $Q(\alpha,\beta)$ can only assume the values $0, 1/2$ or 1. On the other hand there are events α and β such that $Q(\alpha,\beta) \neq 0, 1/2$ or 1; events that are neither perfectly discriminated or equally confused.

We believe that this "paradox" can be resolved by allowing two different scales of discrimination; the conflict arises only because we have at the same time perfect discrimination of utilities and less than total discrimination of probabilities. In (1.6) and (1.7) it is assumed that the ε is the same as in (1.4) and (1.5). It is not obvious that a single ε for both types of discrimination is needed.

The theory of the article [41] is not so much a foundation of imprecise probabilities as an attempt to show that certain reasonable demands on any such theory, embodied in the 11 axioms, is impossible to achieve. Such an impossibility result would have far reaching consequences for decision analysis. We are however not convinced that the axioms really are impossible to satisfy. Two different types of discrimination constants, ε_p and ε_u, for probabilities and utilities, respectively, would seem to resolve the paradox, and we can see no reason why a single discrimination constant for both probabilities and utilities would be necessary in a theory for imprecise representation of probabilities and utilities.

1.3.4 Global belief distributions

One way to introduce second-order probabilities is to use probability distributions on the probabilities of consequences. Second-order probabilities of this type as used in decision analysis are considered in [18]. This model is identical with that proposed in Section 1.4, page 30 and following, and papers III, IV and V modulo terminology and decision rule. With global belief distributions, as opposed to some of the other models considered in this paper, such as [21, 41], not only probabilities but also utilities are allowed to have second-order distributions. In this model every point in the probability space and every point in the utility space is assigned a probability.

As in all theories of imprecise probability expected utility needs to be redefined. The corresponding *generalised* expected utility presented in [18], further described below on page 26 can be easily computed as the scalar product of the mass centres of the probability- and value belief distributions; in a sense this is the standard expected utility with probabilities and values chosen that are the most central in their distributions.

A *global belief distribution* is a probability distribution over a unit cube $[0,1]^n$, giving the distribution of vectors $x = (x_1, x_2, \ldots, x_n)$. For purposes of decision analysis, x is either a vector of the utility values of the n possible consequences of a decision alternative, or a vector of the n probabilities of the consequences. The global distribution $g(p)$ can be interpreted as the second-order probability that the probabilities of the consequences actually is p.

Regarding first-order probabilities, it is in the nature of the problem that we can not compute them exactly, but it would be reasonable to consider the *expected* probabilities.

Assume that we have a global belief distribution g defined on a unit cube B, then we define the expected value of a function of the first-order probabilities $x = (x_1, x_2, \ldots, x_n)$:

Definition 2 $\mathrm{Ep}(f(x)) = \int_B f(x)g(x)\,dV_B(x)$, *where* f *is any function of* x

Then it is shown in [18] that Ep satisfies linearity of expectation.

Let us abuse notation somewhat and call the expected probability of an event A_i with unknown first order probability x_i $\mathrm{Ep}(A_i)$. Then the expected probabilities of an event and a conjunction of independent events follow from Definition 2:

$$\mathrm{Ep}(A_i) = \int_B x_i g(x)\,dV_B(x)\,,$$

i.e. $\mathrm{Ep}(A_i)$ is the moment of x_i,

$$\mathrm{Ep}(\bigwedge_{i=1}^{n} A_i) = \int_B \Pi_{i=1}^{n} x_i g(x)\,dV_B(x)\,.$$

Local distributions

In an even moderately complicated decision situation it may be infeasible for a decision maker to model the probabilities and utilities by global distributions. It is somewhat more realistic to assume that a decision maker can state his or her beliefs through *local* distributions, that is, belief distributions on a particular utility u_i or probability p_i for every i. In the parlance of mathematical statistics the local distributions would be called marginal distributions.

There is a theorem in [18] stating that the global distribution over $x = (x_1, x_2, \ldots, x_n)$ is the product of the local distributions $f(x_i)$. But since the argument rests on independence of the random variables, this can hold only for distributions of utility. First-order probabilities can never be independent random variables since they have to sum to one. In paper V, there is displayed a family of distributions on probabilities that indeed equals the products of their own marginal (local) distributions, but also with an unavoidable normalising multiplicative constant not equal to one. This would in all reason be as close to constructing a global second-order by multiplying local distributions as it is possible to come given the normalisation constraint $\sum_{i=1}^{n} x_i = 1$ for probabilities x_i. That is, multiply the local distributions *and* the normalising constant.

Generalised Expected utility

The goal in this type of decision analysis is to maximise the expected utility $E(C_i) = \sum_{j=1}^{m_i} p_{ij} v_{ij}$ but what values of p_{ij} and v_{ij} should be used when probabilities and utilities are described by belief distributions?

Decision scenarios and *generalised expected utility* are concepts that are used in [18] for dealing with belief distributions in decision analysis.

A *decision scenario* is a structure $(D, P, V, \{\mathbf{p}_i\}_{i=1,\ldots,n}, \{\mathbf{v}_i\}_{i=1,\ldots,n})$, where

- D is a decision situation $\{\{c_{ij}\}_{j=1,\ldots,m_i}\}_{i=1,\ldots,n}$.
- $P = (p_{11}, p_{12}, \ldots, p_{1m_1}, \ldots, p_{n1}, \ldots, p_{nm_n}$ is a unit cube.
- $V = (v_{11}, v_{12}, \ldots, v_{1m_1}, \ldots, v_{n1}, \ldots, v_{nm_n}$ is a unit cube.
- $\mathbf{p}_i$ is a global belief distribution over the unit cube $P_i = (p_{i1}, \ldots, p_{im_i})$ such that $\mathbf{p}_i(x) = 0$ when $\sum_{j=1}^{m_i} p_{ij} \neq 1$.
- $\mathbf{v}_i$ is a global belief distribution over the unit cube $V_i = (v_{i1}, \ldots, v_{im_i})$.

Definition 3 *[18] Let* $(\{C_i\}_{i=1,\ldots,n}, P, V, \{\mathbf{p}_i\}_{i=1,\ldots,n}, \{\mathbf{v}_i\}_{i=1,\ldots,n})$ *be a decision scenario. Then the* generalised expected utility for C_i *is*

$$\int_{P_i \times V_i} \left(\sum_{j=1}^{m_i} x_{ij} y_{ij} \right) \mathbf{p}_i(x_{i1}, \ldots, x_{im_i}) \mathbf{v}_i(y_{i1}, \ldots, y_{im_i}) \, dV(x_{i1}, \ldots, x_{im_i}) \, dV(y_{i1}, \ldots, y_{im_i}) \, .$$

Obviously, the generalised expected utility as defined above is not easy to calculate, but it becomes less cumbersome with the use of the *centroids* of the global belief distributions.

Centroids

Definition 4 *Given a unit cube* $B = (b_1, \ldots, b_k)$ *and a global belief distribution* $g_B \in \mathrm{GBD}(B)$ *the* centroid *of* g_B *is the point* $x_{g_B} = (\beta_1, \ldots \beta_k)$ *in B whose i:th component is*

$$\beta_i = \int_B x_i g_B(x) \, dV_B(x).$$

That is, the centroid consists of the expected probabilities of $\mathrm{Ep}(x_i)$.

And correspondingly for local belief distributions:

Definition 5 *Given a unit cube* $B = (b_1, \ldots, b_k)$ *and a local belief distribution* $f_{B_i} \in \mathrm{LBD}(b_i)$ *the* centroid *of* f_{b_i} *is the point in* b_i *defined by*

$$x_{f_{b_i}} = \int_{b_i} x_i f_{b_i}(x_i) \, dV_{b_i}(x_i).$$

The generalised expected utility can be calculated as the inner product of the centroids of the global belief distributions of probabilities and utilities, this is stated more carefully in the following theorem, proved in [18].

Theorem 1 *If* $(D, P, V, \{\mathbf{p}_i\}_{i=1,\ldots,n}, \{\mathbf{v}_i\}_{i=1,\ldots,n})$ *is a decision scenario then the generalised expected utility*

$$G(C_i) = \langle x_{\mathbf{p}_i}, x_{\mathbf{v}_i} \rangle \, ,$$

where $x_{\mathbf{p}_i}$ *and* $x_{\mathbf{v}_i}$ *are the centroids of* $\mathbf{p}_i$ *and* $\mathbf{v}_i$*, respectively.*

Ranking decision alternatives by generalised expected utility as the inner product of centroids is as fast and uncomplicated as with expected utility with exact probabilities and utilities. Since some of he information in the distributions is disregarded when using centroids, generalised expected utility can be a crude measure, and the decision process should be supplemented with sensitivity analysis when belief distributions are not very warped and certainly when belief distributions are concave. But when a strong warping effect is present, the discarded information makes little difference and ranking through generalised expected utility is relatively stable. An alternative decision rule that preserves more of the information of the underlying distributions is suggested in paper VI.

1.3.5 Fuzzy probabilities

An approach akin to belief measures is to consider probabilities as fuzzy numbers. In [60] fuzzy optimisation problems where fuzzy numbers are compared via ordering cones are considered. One type of optimisation problem looked at in [60] is problems where the objective function has fuzzy-valued coefficients, that is, to minimise $\tilde{f}(\mathbf{x})$ subject to the constraints $[\tilde{g}_i](\mathbf{x}) \in -\mathscr{C}_c, i = 1, 2, \ldots, m$ and $\mathbf{x} \in \mathbf{R}^n$, where $\tilde{a}$ is a fuzzy number, $[\tilde{a}]$ is an equivalence class induced by a defuzzification function and $-\mathscr{C}_c$ is a convex cone.

Fuzzy probabilities are used for modelling uncertainty, but it may be inconvenient to solve a decision problem directly with fuzzy numbers. But one can transform such a problem to an optimisation problem with real coefficients via a defuzzification function η or a a functional ϕ such that $\phi([\tilde{x}]) = \eta(\tilde{x})$, $[\tilde{g}]_i(\mathbf{x}) \in -\mathscr{C}_c$ if and only if $g_i(\mathbf{x}) = \phi([\tilde{g}_i](\mathbf{x})) \leq 0$. The corresponding problem is then to minimise $f(\mathbf{x})$ subject to $g_i(x) \leq 0, i = 1, 2, \ldots, m, \mathbf{x} \in \mathbb{R}^n$.

It is proven in [60] that if $\mathbf{x}^* \in \mathbb{R}^n$ is an optimal solution to the corresponding real optimisation problem then $\mathbf{x}^*$ is a Pareto optimal solution to the original fuzzy problem.

For our purposes we want to study optimisation problems where the objective function is the expected utility of an action and the probabilities and utilities are fuzzy. In the fuzzy optimisation problem above, only the coefficients are allowed to be fuzzy, but the constraints are also fuzzy, so we could have probabilities as the fuzzy coefficients and utilities as the real-valued variables, or the other way around. E. g. if we want to express that utility u_i is somewhere around a we could add the constraints $\tilde{a} \ominus \tilde{1}u_i \leq \tilde{0}$ and $\tilde{1}u_i \ominus \tilde{a}$, where $\tilde{1}$ has the same shape as $\tilde{0}$.

But in this way all inequality constraints are in effect equalities and there is only one solution, namely the expected utility $\sum_{i=1}^{n} p_i u_i$, where $p_i = \eta(\tilde{p}_i)$ and $u_i = \eta(\tilde{u}_i)$. If the defuzzification function η then is the centre of mass, we have in essence the generalised expected utility of [18], except that probabilities and utilities are fuzzy numbers and do not have to sum to one.

1.3.6 Comparison of models

Our interest here is mainly in second-order models and in the efficiency of decision-making systems, both with regards to faithfully representing uncertainty and with regards to computational complexity.

Of the models considered above, the possibilistic hierarchy of [9] is the most general, giving a foundation on which a wide range of models for uncertainty and methods of ranking decision alternatives can be built. Since the degree to which a decision maker can easily and accurately express his or her knowledge about the problem parameters, and the computational complexity of a system for decision analysis, depends on factors not specified by [9], little can be said of such practical matters. In other words, all details of implementation are left open, even if some alternatives are discussed. The concepts in the theory of [9] are not easily grasped, but such conceptual complexity does not preclude the implementation of an efficient decision analysis system that bridges the potential gap between the advanced theory and the decision maker with a user-friendly interface. This is simply because the underlying theory could be hidden from the user. The possibilistic hierarchy allows for the kind of second-order representations of uncertainty under study in this thesis.

The theory of epistemic reliability, a designated second-order theory, in [21], is of great historical interest, and includes a method, maximin, for ranking decision alternatives. Little is said in [21], though, about how representation of the decision maker's information affects the outcome of the decision process, or how efficiently the decision process may be implemented.

The concept of likelihood discrimination in [41] is of purely theoretical interest, making a questionable claim about the impossibility of second-order modelling of uncertainty. In contrast, the decision analysis concept in [18] utilises the properties of second-order probabilities not only for representing uncertainty but also for reducing computational complexity.

Note that the model of global belief distributions as presented in [18] is identical to the second-order probability model that is the topic of this thesis, particularly Section 1.4. The difference lies partly in terminology and partly in how the structural properties of second-order distributions are utilised to motivate different aspects of second-order probability modelling.

Fuzzy numbers, as in [60], have much in common with the second-order probabilities of [18], but do no have to sum to one. The second-order effects that allow for reduction of computational complexity in [18] and restriction of choice in distributions in this thesis depend on second-order probabilities actually being probabilities, so for these desired effects to occur in a fuzzy context, normalisation would be necessary.

1.4 Second-order Probability

Since probabilities are designed to handle uncertainty to begin with it is natural to consider expressing uncertain probabilities with probabilities. When it is uncertain whether an event will occur it is common to express the probability that it will occur, here there is uncertainty as to the precise value of a probability, the uncertain event is that a probability has a particular value. A *second-order probability distribution* is a multivariate probability distribution on n non-negative variables x_i with the restriction that $\sum_{i=1}^{n} x_i = 1$.

But for a formal definition of second-order probability distributions we need a definition of what a probability distribution is.

Definition 6 *[4] For a k-dimensional random vector* $(X_1,\ldots,X_k)$ *the (joint)* distribution μ *is defined by*

$$\mu(A) = \Pr[(X_1,\ldots,X_k) \in A], A \in \mathscr{R}^k$$

where $\mathscr{R}^k$ *is the* σ*-field generated by the bounded rectangles* $[x = (x_1,\ldots,x_k) : a_i < x_i \leq b_i, i = 1,\ldots,k]$.

Definition 7 *[4] A k-dimensional random vector* $(X_1,\ldots,X_k)$ *and its distribution have* density f *with respect to Lebesgue measure if* f *is a non-negative Borel function on* R^k *and*

$$\mu(A) = \int_A f(\mathbf{x})\, d\mathbf{x}, A \in \mathscr{R}^k.$$

Definition 8 *[4] If the k-dimensional vector* $X = (x_1,\ldots x_k)$ *has distribution* μ *and if* $\pi_j : R^k \to R$ *is defined by* $\pi_j(x_1,\ldots,x_n) = x_j$*, the (univariate)* marginal distributions *of* μ *are* $\mu_j = \mu \circ \pi_j^{-1}$ *given by* $\mu_j(A) = \mu[(x_1,\ldots,x_k) : x_j \in A] = \Pr[X_j \in A]$ *for all* $A \in \mathscr{R}$.

Definition 9 *A* second-order probability distribution *is a distribution* μ *with support on a set* $\mathscr{P} = \{(x_1,\ldots,x_k) : 0 \leq a_i \leq x_i \leq b_i, i = 1,\ldots,k, \sum_{i=1}^{k} \leq 1\}$.

Is there then a need to discriminate between probability values in a hierarchical, second-order fashion? And is such discrimination humanly possible? As is pointed out in [24], humans are rarely able to conceive of such second-order discrimination to any higher degree than to state that e.g. 0.2 might be more likely than 0.5 as a probability value. This fact does little to dilute the seriousness of non-discrimination.

The issue lies to a large extent in probability theory itself. A human might not have the capacity for discrimination of probabilities but the constraint that probabilities of the same outcome space must sum to one and be non-negative causes dependency and gives structure to matters pertaining to probability, first- or second-order. In other words, a person can be indifferent towards different second-order distributions, but probability theory is not, even if I can not discriminate, mathematical reality can. The structural aspects of second-order probability that distinguishes different distributions are pervading this thesis.

Ellsberg's paradox, [20], can be explained as the tendency to bet on events with less ambiguous probability, or as uncertainty aversion. Uncertainty aversion is a special case of the general problem of first-order probabilities not offering enough discrimination and literature that point out the usefulness of second-order probabilities for uncertainty aversion include [7, 31, 22, 43].

For a theory of imprecise probability to be useful in practice it ought not be so complicated that very few specialists are able to comprehend workings of the model. Imprecise probabilities should ideally faithfully reflect a decision maker's beliefs and their mathematical representation ought to be sound. Further, a model for imprecise probabilities need to include efficient and rational decision rules if it is to be of practical value in decision analysis. There is a measure of arbitrariness as to decision rules in most models of imprecise probability, even second-order probability, but at least the second-order model offer decision rules such as those in [18, 6] and paper VI that do not rely on the optimistic or pessimistic world views required by the maximax or minimax decision rules. The demand for simplicity is on the surface not met by second-order probability. This apparent complexity is possibly due to the need for higher mathematics when proving properties about second-order probabilities. However, the idea is as simple as probability itself. Admittedly, this is not much of a claim since the concept of probability is far from straightforward, but anyone that has some concept of what a probability is can also understand second-order probability. In contrast, the traditional theories of imprecise probability are quite challenging to grasp, see for instance [55] or [58], where the theoretical level is high. Seen as a theory of imprecise probability, second-order probability has the further advantage of riding piggy-back on probability theory whereas in other theories such as fuzzy probability [61] or lower previsions [55] the whole apparatus needs to be constructed from scratch.

Criticism of second-order probabilities is expressed in [55], the objections stated there may be summarised in that it will be complicated to construct a precise second-order distribution and that in practice there is considerable arbitrariness in the choice of distribution. Since arbitrariness is a key issue of this work, as suggested by the thesis title, some clarification of the concept is justified. Arbitrariness would in this context mean that given some vague, imprecise or uncertain claims about probability values, there are many different second-order probability distributions that would be consistent with these claims and that there is little or no reason to choose one particular second-order distribution among these. Whether the allegedly equivalent alternatives exist due to conflicting opinions in a group or within the mind of an individual or whether the uncertainty is objective or subjective does not matter here. The main conclusion of this thesis is that when dependencies among first-order probabilities are considered the seemingly vast set of second-order distributions consistent with vague probability assessments diminishes in proportion to the certainty of the dependencies.

It is claimed in [24] that second-order probability distributions can never be sharp in the sense of giving exact second-order probability values to first-order probabilities. Or conversely, that second-order probability only can be expressed in terms of inequalities such as "it is more likely that the probability is 0.4 than that it is 0.6". These are serious objections, but some of the results in this thesis suggest a path to a solution to both of them. If there are rational and sound restrictions on second-order probability distributions that limit the choice of distributions the arbitrariness argued in [55] is diminished or even removed and the inequalities of [24] might indeed lead to precise distributions, the only distributions that are consistent with the inequalities. In fact, paper V of this thesis demonstrates that given a constraint of minimal total correlation between a continuous second-order distribution and its product of marginals, there exists only one distribution given lower bounds on the first-order probabilities. If this constraint on information content is deemed rational, arbitrariness is no more. However it is conceivable that first-order probabilities are correlated through relations other than that all probabilities sum to one, e.g. that even if the values of two probabilities are unknown it is known that they are almost equal. Future research might reveal other constraints that limit the choice of distributions in such decision situations where minimal total correlation is not appropriate.

In [23] we find the suggestion of third and higher order probabilities, but these concepts appear to add to the arbitrariness of higher-order probability reasoning in that yet another dimension is introduced without any particular guidelines as to how far along the higher-order axis one should go. Simply put, though, there seems to be little reason to deal with third-order probabilities, intuitively second-order probabilities appear to give enough discriminatory power, and it seems truly difficult for any agent to understand the meaning of probabilities of probabilities of probabilities. In [24] the author denounces probabilities of higher order than two for practical purposes but finds in the notion of an infinite progression of higher order probability a philosophical rationale for imprecision in lower and upper probabilities. A formal and strict argument for dismissing third- and higher-order probabilities are found in [36].

The aim of this thesis has been to address the apparent difficulties with second-order probability distributions, and primarily the problem of arbitrariness. The results of the thesis indicate that there might be conditions on entropy or correlation among probability variables that limits the choice of second-order distribution to the point of removing arbitrariness. Much work, however, remains to be done. We have yet to discover what conditions on entropy or correlation that are appropriate, and if there is a set of conditions that would apply always or if conditions has to vary with the circumstances. The idea of applying the principle of maximised entropy to decision analysis is not new, see [1]. In [42] it is shown that under special circumstances maximisation of expected utility and minimisation of relative entropy are dual problems. Entropy is also one of the measures used in [5] for evaluating uncertainty in interval-based imprecise probabilities. The articles cited above do not deal directly with the subject matter of this thesis, but their use of information theoretical methods in decision analysis have the potential to shed light on the properties of second-order distributions

Specifically, what is achieved in this thesis may be summarised as follows. Principles of rationality and how an agent's information may be expressed in different models for imprecise probabilities are investigated, see paper I, that second-order effects are present and to be reckoned with also in other models of imprecise probabilities than second-order models is revealed in paper II. Second-order effects, mainly that of *warping*, are studied in the specific case of uniform joint second-order distributions with arbitrary intervals, paper III. Warping could be described as the effect that the variance of distributions, second-order probability distributions or distributions derived from such, e.g. distributions of expected utility, decreases with the number of consequences, all other things being equal. The term warping [17] refers to changes in the shape of probability density function curves. In paper IV distribution of expected utility is given explicit formulation in some special cases where joint second-order distribution is uniform. In paper V a contracted version of the Dirichlet distribution is proven to be the unique family of continuous second-order probability distributions that are normalised products of their own marginal distributions, and in paper VI an example of a decision situation with discrete second-order distributions is given along with a suggestion for a second-order decision rule.

Among what has been written about second-order probability distributions, see for instance [19, 18, 51, 49, 43], with rare exceptions such as [35, 50], there has been little attention to particular second-order probability distributions or their relative merits. But as is pointed out in e.g. [34], how can an agent be expected to assign probabilities to probability values when it is difficult enough to come up with a single probability? Then again, a second-order probability distribution that is chosen more or less arbitrarily as to satisfy some intuitions on the reasonableness of probability values would be prone to induce dependencies between the variables, dependencies other than the obvious one that the variables sum to one. These dependencies may or may not reflect reality, but whether they do or not must be considered if one seeks an accurate model. Might there not be some set of principles that limits the choice of second-order distribution so that the arbitrariness issue raised by e.g. [55] may be alleviated? Different families of distributions have different properties with respect to e.g. entropy and dependence among variables and depending on the nature of the probabilities some distributions might be more suitable than others. If imprecise probabilities are to be modelled with second-order probabilities, the question of which second-order distribution is to be employed in which situation must be addressed.

That is, a set of questions for second-order probability imprecision modelling that is as important as it is neglected is what properties are necessary or at least desirable in second-order probability distributions, do these properties change with different circumstances, and which second-order distributions have these properties? In this thesis we see a step towards the resolution of these issues. For instance, in paper III, the marginal distributions of the uniform joint second-order distribution with arbitrary intervals are computed, in paper IV the distribution of expected utility resulting from the uniform multivariate second-order distribution and uniform distribution on utilities are given explicit formulae. In paper V the unique family of continuous second-order probability distributions such that the joint distribution is the normalised product of its own marginal distributions is found to be a generalisation of a special case of the Dirichlet distribution.

The modified Dirichlet distribution from paper V mentioned above is an example of how application of some information theoretic principle leads to significant restrictions in the choice of distribution. The property that a joint distribution factors into its own marginal distributions with a normalising constant is equivalent to a balance of entropy between joint and marginal distributions, or minimal total correlation among the variables. In paper V it is shown that such a distribution is unique given the number n of possible outcomes and the lower limits of support for each first-order probability variable. It is often held against the idea of using second-order probabilities that it is too hard for an agent to decide the second-order probabilities, but in a case where minimal total correlation of first-order variables is suitable the only parameters that needs to be decided are the lowest possible values of the first-order probabilities. It remains to be seen how frequent such situations are, what other principles that apply in the other cases and whether there are corresponding unique families of distributions.

1.5 Conclusions and Further Research

In this thesis some specific second-order probability distributions are investigated. From the debate on second-order distributions it is possible to get the impression that second-order distributions are exchangeable and that the choice between them is arbitrary or a matter of taste. But when, as is done in this work, particular distributions are considered, it becomes clear that there are fundamental differences between them. A statistician knows what distribution to use according to the nature of available data. Similarly, we ought to be able to choose second-order probability distributions after the manner that our imprecise probability information is gathered. And as important that dependency among variables is in mathematical statistics, attention should be paid to dependency among the variables of second-order distributions.

Different distributions have different properties and in order to make an informed choice as to which second-order distribution to use as representative of some particular imprecise probabilities it is necessary to be aware of the distributions' relevant properties. A decision maker might believe that as long as a second-order distribution assigns probability values that do not conflict with her beliefs about the imprecise probability in question, the distribution is feasible. But the complex interplay between the joint distribution on the global level and the marginal distributions on the local level causes effects that ought to be reckoned with. Also, properties of second-order distributions might be exploited for efficiency in decision support systems. Thus the main purposes of considering such properties are on the one hand motivating efficient decision support systems and on the other hand to reveal that the arbitrariness in choosing second-order probability distributions is only apparent. The warp effect is one such property discussed in this thesis, the distribution of entropy among multivariate (global) and marginal (local) distributions is another.

The significance of the warp effect lies partly in that anyone who wishes to use probability distributions to describe imprecise probabilities should be aware that in the marginal distributions more belief mass is located near the lower end of the feasible interval of probability intervals, even if the joint distributions is uniform, see paper III . Another consequence of the warp effect is that concentration of belief in some cases makes it possible in practice to ignore much of the interval between extreme probability values, even to the point of letting the precise mean value, or centroid in the parlance of global belief distributions as in [18], represent the imprecise probability value. With centroids calculation of generalised expected utility becomes trivial, but also the less extreme case where probability intervals are shortened by removal of unlikely values has potential for efficiency in that the risk of overlapping expected utility intervals is reduced.

The warp effect also inspires the idea that with second-order probability distributions, the restriction that first-order variables are non-negative and sum to one forces a shift of belief and therefore also of information content to occur when taking marginals. For instance, we have seen that a uniform joint distribution has polynomial marginal distributions that are far from uniform. The joint distribution is non-informative, but the marginal distributions are not, the structural information inherent in second-order distributions is lacking in the joint distribution and is passed on to the marginal distributions. When choosing a second-order distribution to express some beliefs about imprecise probability, it might be questioned whether it is resonant with those beliefs that all the information content falls on the marginal distributions and none on the joint distribution. If not, it might be a more desirable property of a second-order distribution that information content, or inversely, entropy, is in a sense evenly divided among on the one hand the joint distribution and on the other hand the marginal distributions. It is shown in paper V of this thesis that there is a unique family of distributions that have this property, and that it is a generalisation of a special case of the Dirichlet distribution. The property of even division of entropy among global and local levels turns out to coincide with the property of a joint second-order probability distribution to be equal to the normalised product of its own marginal distributions. This factoring into marginals in turn suggests a measure of independence among the variables. The first-order variables of a second-order distribution are probabilities of one and the same outcome space, the values sum to one, hence these variables can not be independent. But that a distribution equals a product of marginal distributions, even if normalised by a constant, suggests that dependence is restricted to the necessary minimum.

These results raises new questions. We have seen that the properties of second-order distributions are to be reckoned with, and that consequently choice of second-order distributions is not as arbitrary as it was previously thought to be. But what, precisely are the relevant properties? Entropy and dependency are relevant issues wherever probability is involved, but are there other matters to consider also? And there is still a need to clarify what is meant by appropriate conditions on entropy and correlation, or if there are other relevant measures of dependence. The intuition so far is that either the first-order probabilities have no other dependencies other than what is induced by summing to one, or there are such other dependencies. In the first case, the shifted Dirichlet distribution or its hypothetical discrete counterpart would fit the bill, but what second-order distributions are to be used in an accurate model for imprecision in the other case? And in which situations, if any, do these two different cases occur?

Updating also appears to be relevant for future research. When new information is accrued, the degree of imprecision would lessen and the model should change accordingly. The shifted Dirichlet distribution from paper V has the lower bounds of the first-order probabilities as parameters, so updating would reasonably entail increasing the value of these lower bounds. That is, through an observation one would conclude that a lower bound is higher than one previously thought. For example, for some reason I knew yesterday that the probability of seeing a three headed goat on my way to work on any given morning was at least 0.1. But new information convinces me that the probability is at least 0.3. It seems far fetched indeed to conceive of the experiment that led to that updating. Neither seeing or not seeing a three headed goat leeds to the conclusion that the probability discussed here is larger than any particular value. Note that in a setting different from that of papers V and VI updating does result in continuous cumulative second-order distributions in [37].

Updating of discrete second-order distributions comes more natural, at least if one allows the interpretation of probabilities as relative frequencies. If I see the letter ‘b’ in a book that contains 1 000 characters, I know that the probability of seeing a ‘b’ is at least .001. Updating of discrete second-order distributions is touched upon in paper VI. The question of updating is one reason to pursue discrete second-order distributions. One of the most immediate concern for future research is to characterise the discrete second-order distributions that factor into marginals, i.e. to find the discrete counterpart to the shifted Dirichlet distribution in paper V.

Even if continuous second-order expected utility is easily simulated with e.g. Monte Carlo methods, and if discrete second-order expected utility can be computed efficiently with brute force, it would still be of both theoretical and practical importance to find closed expressions for distributions of expected utility. Such expressions of expected utility can in all likelihood not be found for arbitrary second-order probability distributions. Hence this line of research needs to start from particular families of second-order probability distributions.

2. Contributions

2.1 Cross-Disciplinary Research in Analytic Decision Support Systems

The article I, co-authored to a minor degree by the author of this thesis, considers the importance of decision support systems for efficient and accurate decision analysis and the requirements, theoretical and practical, of the principles behind the algorithms of such support systems. The principles of decision analysis and certain problems such as how the concept of rationality can or can not be formalised and how difficult it is for individuals to distinguish between probabilities are discussed.

Different models of representing imprecise information are compared from a theoretical viewpoint and also with regards to how a decision maker is able to express relevant information. Different principles of ranking decision alternatives are also investigated with respect to computational complexity. The article then proceeds to explain the details of a particular decision support system and gives examples of its use in practice.

The article connects theory and practice in an unusual way. A decision support system is not reliable unless the decision maker's information can be entered into the support system without too much distortion and is of little use if it is not efficient with respect to running time. The importance of the approach in paper I is that both these aspects are considered with regards to the design of a decision support system.

2.2 Structure Information in Decision Trees and Similar Formalisms

Article II firstly makes a brief historical survey of decision analysis models and then continues by suggesting a new paradigm. In the so called *first generation models*, introduced by e.g. [53, 45], probabilities and utilities are assumed to have precise values. In realistic decision problems this reliance on precise values is problematic; rarely can a decision maker be expected to provide precise values for probabilities and utilities. Although these classical models can be expanded to handle imprecision, they are not built for imprecision. Thus, it takes some effort to consider vague and imprecise probabilities and utilities, further, the results in terms of finding the optimal decision are not always in accord with common sense.

Other models that are from the outset designed for imprecision are jointly named *second generation models* in the paper. Some examples are [26, 27, 54]. These second generations models are based on many different concepts, e.g. capacities, belief functions, sets of probability measures or upper and lower probabilities. In either case, the decision maker is by design allowed to use imprecise probabilities and utilities, which is the strength of these models. However, this strength is also the weakness; evaluation of expected utility becomes computationally expensive when probabilities and utilities are not numbers. Also, these models typically relies on some kind of interval estimates of probabilities and utilities, this in turn leads to expected utility being an interval rather than a number and comparison of different alternative's expected utilities sometimes become impossible due to overlapping.

A common theme for the thesis is that second-order distributions are inherently present even if only interval representations are explicit, and that these underlying distributions contain information that the decision maker may not have intended. What the article denotes *third generation models* take these effects, chiefly that of *warping*, most of the mass under a distribution functions being close the centroid, into account.

Therefore paper II is highly relevant to this thesis, it demonstrates that second-order information cannot be ignored. The most original aspect of the paper is the way that second-order decision analysis is motivated by second-order effects. The author of this thesis has participated mainly in the section concerning third generation models.

2.3 Warp Effects on Calculating Interval Probabilities

In this article, paper III, where the author of this thesis is the main author and has produced the calculations behind the results, we look at the warp effect on second-order probability density functions of the probabilities of n possible outcomes under the assumption that the probabilities are restricted by intervals. That is, for $n-1$ probabilities p_i, p_i is at least a_i and at most b_i, the n-th probability $p_n = 1 - \sum_{i=1}^{n-1} p_i$ and it is also assumed that each point in the polytope formed by $p_i \in [a_i, b_i], i = 1, \dots n-1$ and $1 - \sum_{i=1}^{n-1} p_i = p_n$ is equally likely, representing a lack of information other than the interval boundaries about the probabilities.

The second-order probability density function $f(p_1)$ of probability p_1 is in the paper produced by a geometric argument; projecting the above mentioned polytope on the p_1-axes. The density function is a, in general piecewise defined, polynomial function with degree $n-2$. Then,given the probability density function, computing the mean is straightforward.

We note the warp effect namely that as n grows, the more of the area under the graph of the second-order probability density function is concentrated to the lower admissible values of probability p_1. As can be expected, the effect is accentuated when the intervals are wide, e.g. in the extreme case where all lower boundaries are equal to zero and upper boundaries one, but less so when the intervals are narrow. So, stating an interval for probabilities also means defining a second-order distribution; the choice is whether to provide one's own or be satisfied with one induced by the intervals and uniform belief in all probability vectors. Hence the results of paper III have important implications for both interval-based and second-order decision analysis.

2.4 Some Properties of Aggregated Distributions over Expected Values

The topic of the fourth article of the thesis, IV, is the distribution of expected utility, limited to decision trees of depth one, uniform belief over $[0,1]$ in all utilities and uniform joint belief in all probability vectors as in paper III, restricted only by probabilities being non-negative and summing to one. Further, only the cases of two, three and four possible outcomes are considered. These restrictions are for simplicity, the general problem requires more sophisticated methods.

Distribution functions for expected utilities have to our knowledge not been given explicit expression before this. Nevertheless, the problem is important for second-order decision analysis; in the end, decision alternatives are in the utility model discriminated by their expected utility. Thus we would benefit from knowing the strength of the warping effect in different circumstances with regard to e.g the number of consequences in order to determine for which types of decision situations different decision rules such as generalised expected utility [18] should be applied. Also, even though expected utility computations can be made efficiently through simulation as in [12, 6], explicit distribution formulae hard-wired into decision support systems could but increase efficiency.

It can be expected from the law of large numbers that expected utility will tend towards the middle value 1/2 under these conditions, the question here is how fast and how, exactly, the density function of expected utility behaves for different numbers of possible outcomes. We see that the warp effect is present, though modestly, already when there are only two possible outcomes, and clearly stronger when there are three or four possible outcomes.

We produce the distribution of expected utility by two different methods; the first is by geometry (giving results for two and three possible outcomes), similar to the method in paper III but using *box splines* [13], the other is a purely statistical approach (usable also for four possible outcomes), relying on a special case of the probability density function of joint uniform second-order probability from paper III. The density function of expected utility when there are n possible outcomes is a sum of polylogarithms of order $n-1$ and less, multiplied by polynomials of degree up to $n-1$, see [40]. Perhaps surprisingly, expected utility is not beta-distributed. However, the probability density functions for beta distributions with parameters chosen to obtain the same variance as the corresponding density functions for expected utility, are numerically very close (less than 0.01 for all values of expected utility between zero and one) to the density functions for expected utility, at least for two, three or four possible outcomes.

2.5 Shifted Dirichlet Distributions as Second-Order Probability Distributions that Factors into Marginals

The results in paper III about the marginal distributions of the joint uniform second-order distribution being more structured than the joint distribution prompted the question whether the are second-order distributions where the joint and marginal distributions in some sense share structure. This vague intuition was transformed to the idea of balance of information content in joint and marginal distributions and to the very concrete property of a joint multivariate probability distribution to be a product of its own marginal distributions and a normalising constant. If this constant was equal to one, the variables would be independent which is impossible since they are the probabilities of mutually exclusive events.

In paper V, written to a substantial degree by the author of this thesis, it is proved via an argument using convolution and the Laplace transform that a second-order distribution factors into marginals if and only if it is a particular generalisation of a special case of the Dirichlet distribution. Namely, if there are n first-order probabilities, the parameters of the Dirichlet distribution have to be $1/(n-1)$, thus distributions that factors into marginals are a special case of the Dirichlet distribution. On the other hand, a new set of parameters are allowed for, the second-order distributions that factors into marginals can be shifted, or rather contracted, so that the interval of support $[0,1]$ for probability p_i is transformed to $[a_i, 1-\sum_{j\neq i} a_i]$, where a_k is the lowest possible value of probability p_k. The author of this thesis is solely responsible for the mathematical proofs in article V.

As of yet, it is not entirely clear what the full significance of the property of factoring into marginals is. However, given that all second-order probability distributions have an inherent structure in that the variables are probabilities, such distributions can never be completely non-informative. For instance, if as in papers III and IV, the joint distribution is uniform, the structure is as it were forced upon the marginal distributions. For second-order probability distributions, a reasonable application of the principle of maximal entropy [29, 30] might be that the built in structural information is evenly distributed among joint and marginal distributions. Also, in [35], desirable properties of a certain type of Dirichlet distributions are proven. Whether these properties carry over to the distribution family in paper V is a matter for future research.

Another example of the potential usefulness of second-order distribution that factor into marginals is that the kind of relative independence among the first-order variables that is suggested by factoring into marginals might very well correspond to some naturally occurring uncertain events where the probability of one event is independent of that of an other, expect for the fact that all else being equal, if one probability increases by a certain amount, the other must decrease by the same amount.

2.6 Expected Utility from Multinomial Second-order Probability Distributions

Written solely by the author of this thesis, paper VI displays a concrete example of how a discrete second-order probability distribution can be employed in a decision problem. As indicated by paper IV, exact computation of the distribution of expected utility is a problem that is far from being solved in its entirety. In practice, distribution of continuous expected utility is simulated [6, 12].

But given discrete underlying distributions of probability and utility, distribution of expected utility can be computed exactly with reasonable efficiency. Of course the computation time depends on the granularity of the second-order distributions. Note that general formulae for discrete expected utility are not considered here but that with given discrete distributions of probability and utility, expected utility is computed with a finite number of additions.

The distributions of expected utility are then suggested to form the basis of a decision rule. If decision alternative A has discrete expected utility probability density function $h_A(z)$, the probability that alternative A yields expected utility z and correspondingly alternative B has distribution of expected utility B, the probability that A has higher expected utility than B can be computed as $\sum_z \sum_{x<z} h_A(z)h_B(x)$. This probability value is not a decision rule per se, but depending on the application a rule could be set that if the probability is higher than a certain threshold value, A is to preferred over B.

In the case that the threshold value is not met, the decision alternatives are simply not distinguishable in terms of expected utility. This could in turn be due to actual likeness of the alternatives or because not enough effort has been put into eliciting probabilities and utilities. The less informative the distributions of expected utility are, the higher is the probability that $\sum_z \sum_{x<z} h_A(z)h_B(x)$ is close to 1/2. Hence another possible part of such a second-order decision rule is based on the variance of distribution of expected utility. Namely, if the probability that A has higher expected utility than B is

close to 1/2, and if both h_A and h_B has variance less than some threshold value, a decision maker should be indifferent to alternatives A and B.

A frequentist interpretation of second-order probability could then be based on first-order probabilities as the number of actual observations over the number of potential observations. Since the multinomial distribution gives the probability of a certain number of observations of different types, this was a natural choice of discrete second-order probability distribution. Then again, the choice in paper VI of underlying probabilities $1/n$ of the events to be observed could be criticised for arbitrariness, particularly since there are stated lower limits of the first-order probabilities of occurrence. The multinomial distribution is a suitable starting point for a discussion on discrete second-order probability, but the underlying event probabilities need to be examined more carefully. This is however immaterial to the suggested probability-based decision rule. The results of the example would doubtless be different with another choice of second-order distributions and will serve as interesting comparison when corresponding experiments are carried out with distributions that are better supported by theory.

Litteraturförteckning

[1] A. E. Abbas. Entropy methods for joint distributions in decision analysis. *iEEE Transactions on Engineering Management*, 53(1):146–159, 2006.

[2] T. Augustin. On decision making under ambiguous prior and sampling information. In *Proceedings of ISIPTA'01*, 2001.

[3] T. Augustin. On the suboptimality of the generalized bayes rule and robust bayesian procedures from the decision theoretic point of view - a cautionary note on updating imprecise priors. In *Proceedings of ISIPTA'03*, 2003.

[4] P. Billingsley. *Probability and Measure*. John Wiley & Sons, Inc., 1995.

[5] A. Bronevich and A. Lepskiy. Measuring uncertainty with imprecision indices. In *ISIPTA '07 - Proceedings of the Fifth International Symposium on Imprecise Probabilities and Their Applications*, pages 47–56, 2007.

[6] O. Caster and L. Ekenberg. Combining second-order belief distributions with qualitative statements in decision analysis. Submitted for publication, 2010.

[7] S. H. Chew and J. S. Sagi. Small worlds: Modeling attitudes toward sources of uncertainty. *Journal of Economic Theory*, 139(1):1–24, March 2008.

[8] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, pages 5:131–295, 1954.

[9] G. De Cooman and P. Walley. A possibilistic hierarchical model for behaviour under uncertainty. *Theory and Decision 52 (4)*, pages 327–374, 2002.

[10] F. Cozman. A brief introduction to the theory of sets of probability measures. CMU Tech. report CMU-RI-TR 97-24.

[11] M. Danielson and L. Ekenberg. A framework for analysing decisions under risk. *European Journal of Operational Research, Vol. 104, Issue 3*, pages 474–484, 1998.

[12] M. Danielson, L. Ekenberg, and A. Larsson. Distribution of expected utility in decision trees. *International Journal of Approximate Reasoning*, 46(2):387–407, 2007.

[13] C. DeConcini and C. Procesi. The algebra of the box spline. http://www.citebase.org/abstract?id=oai:arXiv.org:math/0602019, 2006.

[14] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics, vol xxxviii*, pages 325–399, 1967.

[15] D. Denneberg. *Non-Additive Measure and Integral.* Kluwer, 1994.

[16] L. Ekenberg. *A Unified Framework for Indeterminate Probabilities and Utilities.* PhD thesis, Stockholm University, Department of Mathematics, 2005.

[17] L. Ekenberg, M. Danielson, and J. Thorbiörnson. Multiplicative properties in evaluation of decision trees. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14(3), 2006.

[18] L. Ekenberg and J. Thorbiörnson. Second-order decision analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 9, No 1*, 9(1):13–38, 2001.

[19] L. Ekenberg, J. Thorbiörnson, and T. Baidya. Value differences using second-order distributions. *International Journal of Approximate Reasoning*, 38(1):81–97, 2005.

[20] D. Ellsberg. Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics, 75*, pages 643–669, 1961.

[21] P. Gärdenfors and N.-E. Sahlin. Decision, probability and utility: Selected readings. In *Decision, Probability and Utility:Selected Readings*, chapter 16, Unreliable probabilities, risk taking, and decision making, pages 313–334. Cambridge University Press, 1988.

[22] R. Giraud. Objective imprecise probabilistic information, second order beliefs and ambiguity aversion: an axiomatization. In *ISIPTA*, pages 183–192, 2005.

[23] I. J. Good. *Good Thinking: The foundations of Probability and Its Applications.* Univ. of Minnesota Press, 1983.

[24] I.J. Good. Subjective probability as the measure of a non-measurable set. In P. Suppes, B. Nagel, and A. Tarski, editors, *Logic, Methodology, and the Philosophy of Science*, pages 319–329. Stanford University Press, 1962.

[25] I.J. Good. Bayesian statistics, vol 1. In D.V. Lindley J. M. Bernardo, M.H. DeGroot and A.F.M. Smith, editors, *Bayesian Statistics, Vol 1*, chapter Some history of the hierarchical Bayesian methodology, pages 489–519. Valencia University Press, 1980.

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

[27] L. Hurwicz. Optimality criteria for decision making under ignorance. *Cowles Comission Discussion Paper 370*, 1951.

[28] J-Y. Jaffray. Rational decision making with imprecise probabilities. In *Proceedings of ISIPTA'99*, 1999.

[29] E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.

[30] E.T. Jaynes. Information theory and statistical mechanics ii. *Physical Review*, 108(2):171–190, 1957.

[31] P. Klibanoff, M. Marinacci, and S. Sukerji. A smooth model for decision making under ambiguity. *Econometria*, 73:1849–1892, 2005.

[32] B. O. Koopman. The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41:269–292, 1940.

[33] B. O. Koopman. The bases of probability. *Bulletin of the American Mathematical Society*, 46:763–774, 1940.

[34] H. E. Kyburg. Higher order probabilities and intervals. *International Journal of Approximate Reasoning 2*, pages 195–209, 1988.

[35] J. Lawry. *Inexact Reasoning*. PhD thesis, University of Manchester, 1994.

[36] J. Lawry. An inductive method for inexact reasoning. *International Journal of Approximate Reasoning*, 16(2):205 – 221, 1997.

[37] J. Lawry, I. González-Rodríguez, and Y. Tang. Imprecise probabilities from imprecise descriptions of real numbers. In *Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 277–286, 2009.

[38] I. Levi. On indeterminate probabilities. *The Journal of Philosophy, vol. 71*, pages 391–418, 1974.

[39] I. Levi. *The Enterprise of Knowledge*. MIT Press, 1980.

[40] L. Lewin. *Polylogarithms and Associated Functions*. North Holland, 1981.

[41] R.D. Luce and H. Raiffa. *Games and Decisions*, chapter Appendix 1, A probabilistic theory of utility, pages 371–384. Dover Publications, 1957.

[42] R. Nau, V. R. R. Jose, and R. Winkler. Duality between maximization of expected utility and minimization of relative entropy when probabilities are imprecise. In *ISIPTA'09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 337–346, 2009.

[43] R. F. Nau. Uncertainty aversion with second-order utilities and probabilities. *Management Science*, 52(1):136–145, 2006.

[44] Y. Pan. *Fuzzy Probability Calculus*. PhD thesis, Binghamton University (SUNY), 1997.

[45] L.J. Savage. *The Foundation of Statistics*. John Wiley and Sons, 2nd edition, 1972.

[46] G. Shafer. *A Mathematical theory of evidence*. Princeton University Press, 1976.

[47] Z. Shapira. *Risk Taking: A Managerial Perspective*. Russel Sage Foundation, 1995.

[48] C. A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series B, xxiii*, pages 1–25, 1961.

[49] L. V. Utkin. Imprecise second-order hierarchical uncertainty model. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Volume 11:3*, pages 301–317, 2003.

[50] L. V. Utkin. Probabilities of judgments provided by unknown experts by using the imprecise dirichlet model. *Risk, Decision and Policy*, 9(4):371–389, January 2004.

[51] L. V. Utkin and T. Augustin. Decision making with imprecise second-order probabilities. In *ISIPTA '03 - Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, pages 547–561, 2003.

[52] T. Denœux, Z. Younes, and F. Abdallah. Representing uncertainty on set-valued variables using belief functions. *Artificial Intelligence*, 174(7-8):479–499, 2010.

[53] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1944.

[54] A. Wald. *Statistical Decision Functions*. John Wiley and Sons, 1950.

[55] P. Walley. *Statistical reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.

[56] P. Walley. Imprecise probabilities. In *Encyclopedia of Statistical Sciences, Update Vol 3.*, pages 355–359. Wiley, New York, 1999.

[57] S. R. Watson, J.J. Weiss, and M. L. Donnell. Fuzzy decision analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, 9:1–9, 1979.

[58] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3):149 – 170, 2000.

[59] P. M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

[60] H.-C. Wu. Fuzzy optimization problems based on ordering cones. *Fuzzy Optimization and Decision Making*, 2 (1):13–29, 2003.

[61] L. A. Zadeh. Fuzzy probabilities. *Information Processing and Management, 20*, pages 363–372, 1984.