Uncertainty Levels of Second-Order Probability

David Sundgren and Alexander Karlsson

Abstract—Since second-order probability distributions assign probabilities to probabilities there is uncertainty on two levels. Although different types of uncertainty have been distinguished before and corresponding measures suggested, the distinction made here between first- and second-order levels of uncertainty has not been considered before. In this paper previously existing measures are considered from the perspective of first- and second-order uncertainty and new measures are introduced. We conclude that the concepts of uncertainty and informativeness needs to be qualified if used in a second-order probability context and suggest that from a certain point of view information can not be minimized, just shifted from one level to another.

Index Terms—Uncertainty, entropy, second-order probability.

I. INTRODUCTION

REASONING under uncertainty is a fundamental problem within artificial intelligence. In this probability is an important tool, but in real life situations there is often uncertainty regarding the probability values themselves. Second-order probability, see e.g. [1], [2], [3], is an hierarchical model of imprecise probability that can be used to model different types of uncertainty regarding first-order probability distributions, e.g., in terms of their quality [4]. Just as in e.g. the possibilistic hierarchy [5], the epistemic reliability model [4] or fuzzy probabilities [6], probability distributions are discriminated by weights. In the case of second-order probability the weights are themselves probabilities. Where there are probability distributions there is uncertainty, with a second-order distribution there is then the uncertainty that comes with the second-order probabilities but also the uncertainty of the first-order probabilities.

Thus it is meaningful to distinguish different types of uncertainty, and in the limits of uncertainty, ignorance and uninformativeness. As is pointed out in [7] ignorance comes in different forms, and E. T. Jaynes wrote in [8] that 'A major thing to be learned in developing this neglected half of probability theory is that the mere unqualified epithet "uninformative" is meaningless.'

A. Levels of Uncertainty in Dempster-Shafer

In the literature on Dempster-Shafer theory [9] there is in e.g. [10] a distinction between two types of uncertainty, *dissonance* and *nonspecificity*. Shannon entropy is in [10]

mentioned as an example of a measure of dissonance but not nonspecificity; beliefs expressed in terms of probability distributions are dissonant. Dissonance pertains to probabilistic uncertainty and in e.g. [11] an entropy like measure for dissonance (or discordance) of the basic assignment functions of [9] is introduced. On the other hand nonspecificity is in [10] described as increasing with the number of alternatives in a decision situation and the Hartley measure is put forward as the appropriate measure of nonspecificity. In [12], the measures for discord and nonspecificity are aggregated into *total uncertainty*, see also [13] for a more recent account of uncertainty measures in evidence theory.

Since second-order probability is not equivalent to Dempster-Shafer theory the uncertainty measures designed for belief functions are not directly applicaple to second-order distributions. Yet, in [14] there is a discussion of how second-order distributions could be interpreted in terms of nonspecificty. Smithson [14] recounts the situation in [15] where Miss Julie is invited to bet on the outcomes of three tennis matches in terms of second-order probability. In match A it is known that it will be an even match; in terms of first- and second-order uncertainty there is no second-order uncertainty and maximum first-order uncertainty in match A. Nothing is known about match B and Smithson [14] suggests that the second-order distribution be a uniform distribution spanning the $[0, 1]$ interval, in this case both first- and second-order uncertainty is high but it is questionable whether any second-order distribution can model the ignorance regarding match B. As regards to match C Miss Julie knows that one of the players is excellent and the other an amateur but she does not know which one is the better player, in this case there is no first-order uncertainty but maximum second-order uncertainty. The corresponding second-order distributions could be described as follows in the terms used in the sequel of this paper: In match A we say that the second-order distribution is determined by $p(x_1 = 0.5, x_2 = 0.5) = 1$, in match C we could have the second-order distribution $p(x_1 = 1, x_2 = 0) = 0.5, p(x_1 = 0, x_2 = 1) = 0.5$. The case of match B is probably not possible to fully specify with second-order distributions.

According to Smithson [14] B is the most nonspecific situation, C is more specific than B and A is considered to be "quite specific". With respect to B there are two different approaches; either use a uniform second-order distribution as advised in [14] or refuse to express the uncertainty of B with a second-order distribution. In the first case there is some first-order uncertainty since there is positive probability for high-entropy points. Inasmuch the second-order distribution

assigns belief to points that are far apart there is a high second-order uncertainty but belief in neighboring points reduce second-order uncertainty to be less than maximal.

Comparing nonspecificity with first- and second-order uncertainty, if match B after all is represented by a second-order distribution B would be between A and C along both the first- and second-order uncertainty scale while being the most nonspecific, indicating that first-/second-order uncertainty and dissonance/nonspecifity are independent measures. If on the other hand B is left out for being impossible to be modelled with a second-order distribution, dissonance would have positive correlation to first-order uncertainty.

Even though we discuss uncertainty measures for second-order probability rather than for Dempster-Shafer theory there would be a parallell in that what we call first-order uncertainty could be seen as probabilistic just as dissonance is. Second-order level uncertainty could correspondingly described as deterministic.

B. Probabilistic and Deterministic Uncertainty

The distinction might be clarified with an example. Say that we want to express ignorance as to the outcome of an experiment with a second-order probability distribution. One possibility is to assign all second-order probability to the maximum entropy distribution where all outcomes are equally probable. This way we would express ignorance on what could be called the first-order level. But on the second-order level we are absolutely certain, there is no doubt which first-order distribution is the proper one, all uncertainty if placed on the level of first-order probabilities. This situation might be described as being certain of being uncertain. On the other hand we could express ignorance by the uniform second-order distribution on the zero entropy distributions where one of the outcomes is certain. In a first-order perspective there is no uncertainty, distributions with positive entropy are not considered, however we would know no more about the outcome of the experiment. The uncertainty remains but now entirely on the second-order level, you might say that in this case we are uncertain of being certain.

To make the experiment tangible, consider throwing a die. The second-order distribution mentioned first, where all second-order probability is assigned to the uniform distribution $(1/6, 1/6, \ldots, 1/6)$ could be employed to express certainty in that the die is fair. But if a uniform second-order distribution is put on the six zero-entropy distributions $(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ this would mean that we know for certain that the die is fixed to always show the same number but we have no idea which one. A possible interpretation in the realm of philosophy or psychology could be that the first type of second-order distribution could be used by someone who believes in the fundamental randomness of everything but the other type is suitable for an ignorant[1] determinist. These two

second-order probability distributions are extreme examples and other second-order distributions could represent mixtures of the two types of ignorance. The point is that ignorance is not enough to specify a second-order distribution unambiguously. And even given the distribution of uncertainty between the two levels it is not obvious how to measure uncertainty.

Mork [16] has performed an extensive study on how uncertainty could be measured from credal sets and from second-order probability distributions. He introduced an entropy based uncertainty measure called GSU (Gärdenfors-Sahlin uncertainty) after [4]. Entropy has also been applied to interval-based imprecise probabilities (i.e., without second-order information) [17]. In this paper we will use entropy as a basis for a majority of our uncertainty measures. By using simple numerical examples, we will contrast the result of our measures to previous measures found in the literature.

In particular, we show that uncertainty measures for second-order probability distributions can be constructed in various ways, and that there are seemingly reasonable requirements on uncertainty measures that are not always met for the measures we discuss. We suggest that uncertainty in a second-order probability setting needs to be qualified in order to be measured consistently. There are likely many aspects with which to specify what is meant by uncertainty for second-order distributions, but whether uncertainty is measured on the first- or second-order level, or both, appears to be a relevant specification. To our knowledge the only previous uncertainty measure for second-order probability that could be said to cover both levels is the GSU measure of [16]. We here introduce a measure for aggregated uncertainty that decomposes naturally into first and second-order levels.

C. Definitions and Notation

Let the outcome space Ω have a finite number of elements n, $\Omega = \{s_i : i = 1, \ldots, n\}$. What we call first-order probabilities are the probabilities of the n outcomes and second-order probability distributions are probability distributions with first-order distributions as random variables. That is, any probability can be seen as a first-order probability, but second-order probability is probabilities over probabilities. Since we are more interested in the probabilities of outcomes rather than the outcomes themselves, we denote the probability $\Pr(s_i)$ by x_i, i.e. all x_i are first-order probability values, $0 \leq x_i \leq 1, \sum_{i=1}^{n} x_i = 1$. All marginals mentioned below are one-dimensional marginal probability distributions.

Further, to simplify computations we restrict the first-order probability values to rational numbers, $x_i = k_i/N$, where $\sum_{i=1}^{n} k_i = N$. A probability distribution will then be considered as a vector $\mathbf{x} = (x_1, \ldots, x_n) = (k_1/N, \ldots, k_n/N)$. Let X denote a set of first-order probabilities $\mathbf{x}$ where the marginals are rational numbers.

---

[1]The term is of course not used in a derogatory sense.

Furthermore, let us define the set:

$$X_N = \left\{ \mathbf{x} : \mathbf{x} = (k_1/N, \ldots, k_n/N), \sum_{i=1}^{n} k_i = N, k_i \in \mathbb{N} \right\}$$

i.e., the set of all first-order probabilities in the form of rational numbers that fulfill $\sum_{i=1}^{n} k_i = N$. For every N, X_N is finite so we are therefore restricted to discrete second-order probability distributions over first-order distributions. That is, the discrete second-order probability distributions discussed here have X_N as outcome space. Second-order probability distributions are denoted p, and the probability of first-order probability distribution $\mathbf{x} \in X_N$ is then $p(\mathbf{x})$. The marginal probability of first-order probability value x_i is written $p_i(x_i)$. We remind of the definition of Shannon entropy

$$H(\mathbf{x}) = -\sum_{i=1}^{n} x_i \log_2 x_i \,.$$

II. First-Order Uncertainty

As we have argued we may distinguish two levels of uncertainty. The term first-order uncertainty is intended to capture uncertainty on the level of first-order probabilities in the context of a second-order probability distribution. In other words, first-order uncertainty is the type of uncertainty expressed by probababilies of unknown outcomes of an event. In the absence of second-order probability an uncertainty measure such as entropy would be used, but here we have second-order probability distributions to account for. In this section we present one way of measuring such first-order uncertainty.

Weighted entropy. This uncertainty measure intends to capture the collected amount of entropy in the first-order probability values. Since a second-order distribution assigns probability values to the first-order variables, the entropy values are weighted accordingly. That is, the more probable a vector $\mathbf{x}$ is, the more weight we give to its entropy.

$$W_H(X, p) = \sum_{\mathbf{x} \in X} p(\mathbf{x}) H(\mathbf{x}) = -\sum_{\mathbf{x} \in X} p(\mathbf{x}) \sum_{i=1}^{n} x_i \log_2 x_i \,.$$

III. Second-order Uncertainty

These measures are meant give the degree of uncertainty on the second-order level. For instance, if all second-order probability is given to low entropy points there could still be a high degree of uncertainty in that there is little or no commitment to any one particular outcome.

We have found three basic approaches to second-order uncertainty, one is the entropy of the second-order probability distribution, the other one is based on how much the second-order distribution is spread out. The latter, distance-based measures are justified by the intuition that uncertainty could be expressed by conflicting statements. The third measure considers the volume of the support of the second-order distribution.

Entropy. This measure is simply the entropy $H(p)$ of the second-order distribution:

$$H(X, p) = -\sum_{\mathbf{x} \in X} p(\mathbf{x}) \log_2 p(\mathbf{x}) \,.$$

Weighted Kullback-Leibler divergence. If we want to be able to compare second- and first-order uncertainty, we have to use an entropy-based distance measure so that first- and second-order uncertainty are measured in the same units. If second-order uncertainty is linked to the spread of belief over the probability simplex and measure entropy, we suggest the average Kullback-Leibler divergence, see [18], to the mean as measure of second-order uncertainty, i.e., the mean of the second-order distribution $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$, defined by $\mu_i = \sum_{\mathbf{x} \in X_N} p(\mathbf{x}) x_i$ is used as point of reference.

$$W_{D_{\mathrm{KL}}}(X, p) = \sum_{\mathbf{x} \in X} p(\mathbf{x}) D_{\mathrm{KL}}(\mathbf{x} | \boldsymbol{\mu}) =$$

$$\sum_{\mathbf{x} \in X} p(\mathbf{x}) \sum_{i=1}^{n} x_i \log_2 \frac{x_i}{\mu_i} \,.$$

Degree of imprecision. The degree of imprecision [19] aims to be an approximation of the hypervolume spanned by the first-order probability distributions with positive second-order probability.

$$DI(X, p) = \frac{1}{n} \sum_{i=1}^{n} \left[\max_{\mathbf{x} \in \hat{X}_N} x_i - \min_{\mathbf{x} \in \hat{X}_N} x_i \right] \,,$$

where $\hat{X}_N = \{\mathbf{x} \in X | p(\mathbf{x}) > 0\}$. Note that the computation of DI can be performed by maximizing respective minimizing over the extreme points of the convex hull of first-order probability distributions with positive second-order probability.

IV. Aggregated Uncertainty

Under this rubric we collect measures that could claim to express the aggregated degree of uncertainty on first and second-order level.

Mork's GSU. The uncertainty measure introduced in [16] is named after Gärdenfors and Sahlin and is inspired by their discussion in [4] about epistemic reliability. The idea might be summarized in that the information value is reduced by adding second-order probability through a convex combination of information values.

The measure is defined by:

$$GSU(X, p) = \max_{\mathbf{x} \in X} \left\{ -\sum_{i=1}^{n} \left[x_i \sum_{\mathbf{x} \in X} p(\mathbf{x}) \log_2 x_i \right] \right\} \,.$$

Sum of W_H and $W_{D_{\mathrm{KL}}}$. Since both the first-order measure of weighted entropy and the second-order measure of weighted Kullback-Leibler divergence are based on entropy of first-order probabilities and have the same units the sum of the measures is meaningful.

TABLE I
NUMERICAL EXAMPLES

Examples	First- and Second-order Probability Distributions
A	$X = \{(1/3, 1/3, 1/3)\}, p(1/3, 1/3, 1/3) = 1$
B	$X = \{(32, 29, 29)/90, (29, 32, 29)/90, (29, 29, 32)/90\},$ $\forall \mathbf{x} \in X, p(\mathbf{x}) = 1/3$
C	$X = X_{10}, (\forall \mathbf{x} \in X)(Pol(\mathbf{x}|(1/3, 1/3, 1/3)))$ (Perks' prior)
D	$X = X_{10}, (\forall \mathbf{x} \in X)(Pol(\mathbf{x}|(1/2, 1/2, 1/2)))$ (Jeffreys' prior)
E	$X = X_{10}, (\forall \mathbf{x} \in X)(Pol(\mathbf{x}|(1, 1, 1)))$ (Bayes-Laplace's prior)
F	$X = \{(1/6, 1/6, 2/3), (1/6, 2/3, 1/6)\},$ $p(1/6, 1/6, 2/3) = 3/4, p(1/6, 2/3, 1/6) = 1/4$
G	$Y = \{(2/3, 1/6, 1/6), (1/3, 1/3, 1/3), (1/4, 1/4, 1/2)\},$ $\forall \mathbf{y} \in Y, p(\mathbf{y}) = 1/3$
H	$(X, Y) = X \times Y = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \in X, \mathbf{y} \in Y\},$ $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}), p(\mathbf{x}), p(\mathbf{y})$ from F and G.
I	$(X, Y) = X \times Y = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \in X, \mathbf{y} \in Y\},$ $\forall (\mathbf{x}, \mathbf{y}) \in X \times Y p(\mathbf{x}, \mathbf{y})) = 1/6$, $p(\mathbf{x}), p(\mathbf{y})$ from F and G.
J	$X = \{(1/6, 1/6, 2/3), (1/6, 2/3, 1/6), (7/10, 1/10, 1/5)\},$ $p(\mathbf{x_1}) = 1/2, p(\mathbf{x_2}) = 1/3, p(\mathbf{x_3}) = 1/6.$
K	X as in J, but $\forall \mathbf{x} \in X, p(\mathbf{x}) = 1/3.$

Interestingly, the sum of W_H and $W_{D_{\mathrm{KL}}}$ equals the entropy of the second-order distribution's mean:

Theorem 1: $W_H(X, p) + W_{D_{\mathrm{KL}}}(X, p) = H(\boldsymbol{\mu})$, where $\mu_i = \sum_{\mathbf{x} \in X} p(\mathbf{x}) x_i$.

Proof:

$$W_H(X, p) + W_{D_{\mathrm{KL}}}(X, p) =$$

$$\sum_{\mathbf{x} \in X} p(\mathbf{x}) \sum_{i=1}^{n} x_i \left(\log_2 \frac{x_i}{\mu_i} - \log_2 x_i \right) =$$

$$- \sum_{\mathbf{x} \in X} p(\mathbf{x}) \sum_{i=1}^{n} x_i \log_2 \mu_i =$$

$$- \sum_{i=1}^{n} \left(\sum_{\mathbf{x} \in X} p(\mathbf{x}) x_i \right) \log_2 \mu_i = - \sum_{i=1}^{n} \mu_i \log_2 \mu_i = H(\boldsymbol{\mu})$$

∎

V. NUMERICAL EXAMPLES

To better understand these uncertainty measures we have applied them to some second-order probability distributions with various intuitive uncertainty properties. Some of the distributions have only a few points, other distributions have support on the entire space X_N, for $N = 10$. The latter distributions will come from the multivariate Pólya family [20]:

$$Pol(\mathbf{k}|\boldsymbol{\alpha}) = \frac{N! \Gamma\left(\sum_{i=1}^{n} \alpha_i \right)}{\Gamma\left(N + \sum_{i=1}^{n} \alpha_i \right)} \prod_{i=1}^{n} \frac{\Gamma(k_i + \alpha_i)}{k_i! \Gamma(\alpha_i)},$$

where $\sum_{i=1}^{n} k_i = N$ and α_i are parameters of the corresponding Dirichlet distribution. Note that we can obtain a distribution over X_N by $Pol(\mathbf{k}/N | \boldsymbol{\alpha})$. The Pólya family of distributions can be seen as the discrete counterpart of the Dirichlet family and is the result of integrating out the underlying probabilities drawn from a Dirichlet distribution in a multinomial distribution.

Consider the examples defined by Table I (where $n = 3$) and the results of applying the uncertainty measures, shown in Table II.

The purpose of examples A and B is to how different measures deal with two distributions that put all second-order probability on high entropy points; on the one hand A where the second-order distribution has support only on the maximum entropy point, on the other B where there is a uniform distribution on three different points that are close to the maximum entropy. In Table II we see that H, the entropy of the second-order distribution is the only measure that makes much of the fact that the second-order distribution has support on three points rather than one.

In examples C, D and E we look at symmetric Pólya distributions with parameters $1/3$ (Perks), $1/2$ (Jeffreys) and 1 (Bayes-Laplace), respectively. The first-order level measure W_H and the second-order level measure H increase with the Dirichlet parameters while the other second-order level measure $W_{D_{\mathrm{KL}}}$ decreases. If we interpret the Dirichlet parameters as a measure of the amount of available data higher parameter values would give more (first order-) probabilistic credibility to the mean probability vector.

If the parameters are equal, the mean would be the maximum entropy point of the simplex. In our examples then, first-order uncertainty would increase (as does W_H), and second-order uncertainty decrease (since there is more data to support a particular first-order probability).

We may also note that GSU gives infinite values in these cases. This is because there are zero-valued first-order probabilities with positive second-order probability. It could be argued that zero-valued first-order probabilities x_i should be excluded, since the event that x_i is the probability of is impossible and does not belong in the outcome space. On the other hand it is feasible that (on a second-order level) it is possible but not certain that an event can not occur, i.e. that neither $x_i = 0$ or $x_i > 0$ can be ruled out.

TABLE II
RESULTS OF APPLYING THE UNCERTAINTY MEASURES TO THE EXAMPLES

	First-order	Second-order			Aggregated	
Examples	W_H	DI	H	$W_{D_{\mathrm{KL}}}$	GSU	$H(\mu)$
A	1.5850	0.0000	0.0000	0.0000	1.5850	1.5850
B	1.5834	0.0333	1.5850	0.0016	1.5865	1.5850
C	0.6852	0.5682	5.5146	0.8997	∞	1.5850
D	0.8285	0.5682	5.8399	0.7565	∞	1.5850
E	1.0486	0.5682	6.0444	0.5364	∞	1.5850
F	1.2516	0.5000	0.8113	0.1768	2.0016	1.4284
G	1.4455	0.2500	1.5850	0.1091	1.7233	1.5546
H	2.6972	0.5000	2.3962	0.2858	3.7249	2.9830
I	2.6972	0.5000	2.5850	0.3408	3.4749	3.0379
J	1.2358	0.5333	1.4591	0.3189	2.0803	1.5547
K	1.2200	0.5333	1.5850	0.3633	2.0635	1.5833

VI. PROPERTIES

In [16] there are sets of requirements for uncertainty measures both for credal sets and for second-order probability distributions. Likewise there is in [21] a list of requirements for measures of uncertainty, but designed for belief functions (see [9]). Since the requirements of [21] do make sense when translated to a second-order probability setting we will consider also these, some of them also coincide with the requirements of [16]. Please note that since belief functions in the language of second-order probability is best translated as lower bounds of first-order probabilities that in turn do not determine unique second-order distributions, the translation of properties must at times be ad hoc. Strictly speaking then, the requirements below that are taken from [21] is to be considered as inspired by [21] rather than literal translations.

Below we describe these requirements briefly, as far as they can be expressed in the terms used in this paper. The original authors have different notation, and in the case of [21] information is carried by belief functions instead of second-order distributions. Let $\mathcal{U}(X,p)$ denote an uncertainty measure for a second-order probability distribution p with support on X.

(i) *Conicides with entropy* C1 in [16], (1) in [21]. Uncertainty coincides with entropy if all second-order probability is put on a single vector. That is, If $p(\mathbf{x}) = 1$ for some $\mathbf{x} \in X$, then $\mathcal{U}(X,p) = H(\mathbf{x})$.

(ii) *Continuous* C2 in [16]. $\mathcal{U}$ is continuous in p.

(iii) *Symmetric* C3 in [16]. $\mathcal{U}$ is symmetric, i.e. invariant under permutations in the vectors $\mathbf{x}$, i.e.. if $Y = \{(x_{\pi(1)}, x_{\pi(2)}, \ldots x_{\pi(n)})|\mathbf{x} \in X\}$ where π is a permutation of $\{1, 2, \ldots, n\}$, $\mathcal{U}(X,p) = \mathcal{U}(Y,p)$.

(iv) *Hartley* (2) in [21]. For a uniform second-order distribution, i.e. p s.t. $p(\mathbf{x}) = 1/|X|$ for $\mathbf{x} \in X$, uncertainty equals $\mathcal{U}(X,p) = \log_2 |X|$, the Hartley measure of X.

(v) *Range* (3) in [21]. The range of $\mathcal{U}$ is the interval $[0, \log_2 |\Omega|] = [0, \log_2 n]$.

(v') With an alternative interpretation, (v)' requires the range to be $[0, \log_2 |X|]$.

(vi) *Additive* C4 in [16], (5) in [21]. Additivity, i.e. if X is the cartesian product $A \times B$ and A and B are independent so that $p((\mathbf{x}, \mathbf{y})) = p(\mathbf{x})p(\mathbf{y})$, then $\mathcal{U}(X,p) = \mathcal{U}(A,p) + \mathcal{U}(B,p)$.

(vii) *Subadditive* C5 in [16], (4) in [21]. Subadditivity, i.e. if X is the cartesian product $A \times B$, then

(viii) *Bounded by entropy* NC1' in [16]. The uncertainty of a second-order distribution is at least as high as the entropy of any of the first-order probability distributions in its support. $\mathcal{U}(X,p) \geq \max_{\mathbf{x} \in X} H(\mathbf{x})$.

(ix) *Bounded by credal set* NC3 in [16]. If Π is a partition of X and $\mathrm{Conv}(X)$ is the convex hull of X, then $\mathcal{U}'(\Pi(\mathrm{Conv}(X))) \geq \mathcal{U}(\Pi(X),p)$, where $\mathcal{U}'$ is the corresponding uncertainty measure for a credal set.

(ix') For some of the measures considered here $(W_H, H, W_{D_{KL}}, H(\mu))$ it is not possible to remove the second-order distribution p. But we might formulate a version of NC3 that retains some of what we believe is intended by the requirement. Short of removing p we replace p with the maximum entropy second-order distribution. We declare requirement (ix)' to be that $\mathcal{U}(X,p) \leq \mathcal{U}(X,q)$, where q is the uniform second-order distribution on X.

In Table III we summarize our findings of whether the measures studied here fulfill the requirements. $H(\mu)$ in the rightmost column refers to the sum of W_H and $W_{D_{\mathrm{KL}}}$.

VII. SUMMARY AND CONCLUSIONS

In this paper, we have suggested a division of uncertainty between two levels corresponding to first and second-order probability. With such a division it becomes possible to distinguish between on the one hand uninformativeness in the guise of a uniform probability distribution over the possible outcomes and on the other hand uninformativeness in the form of a uniform second-order distribution. In the first case we are sure of being uncertain but in the other we express uncertainty on a higher level.

We studied the behavior of six different uncertainty measures, two for each uncertainty level and two for

TABLE III

PROPERTIES OF THE UNCERTAINTY MEASURES. NOTES: [1] $W_H + W_{D_{KL}}$ FULFILLS REQUIREMENT (IV) IF AND ONLY IF $\mu_i = 1/n$ FOR $i = 1, \ldots, n$. [2] EXAMPLE A IN TABLE II SERVES AS COUNTER EXAMPLE IN THE NEGATIVE CASES. [3] PROVED IN [16]. [4] EXAMPLES F, G AND H IN TABLE II SERVE TOGETHER AS COUNTER EXAMPLE SINCE THE UNCERTAINTIES OF ROW F AND G SHOULD ADD TO THE VALUE IN H. [5] WE HAVE NOT FOUND NEITHER PROOF NOR COUNTEREXAMPLE. [6] EXAMPLES F, G AND I IN TABLE II SERVE TOGETHER AS COUNTER EXAMPLE SINCE THE SUM OF THE UNCERTAINTIES OF ROW F AND G SHOULD NOT BE LESS THAN THE VALUE IN I. [7] NOT APPLICABLE SINCE WE HAVE FOUND NO WAY OF DEFINING A CORRESPONDING MEASURE FOR A CREDAL SET, I.E. WITH OUT THE SECOND-ORDER PROBABILITY DISTRIBUTION. [8] SEE EXAMPLES J AND K.

		First-order	Second-order			Aggregated	
	Requirement	W_H	DI	H	$W_{D_{KL}}$	GSU	$H(\boldsymbol{\mu})$
(i)	*Coincides with entropy*	Yes	No	No	No	Yes	Yes
(ii)	*Continuous*	Yes	Yes	Yes	Yes	Yes[3]	Yes
(iii)	*Symmetric*	Yes	Yes	Yes	Yes	Yes[3]	Yes
(iv)	*Hartley*	No	No	Yes	No	No	No[1]
(v)	*Range*	Yes	Yes	No	Yes	No	Yes
(v′)		No[2]	No[2]	Yes	No[2]	Yes	No[2]
(vi)	*Additive*	Yes	No[4]	Yes	Yes	Yes[3]	Yes
(vii)	*Subadditive*	?[5]	?[5]	No[6]	No[6]	?[5]	No[6]
(viii)	*Bounded by entropy*	No	No	No	No	Yes[3]	No
(ix)	*Bounded by credal set*	N. a.[7]	Yes	N. a.[7]	N. a.[7]	Yes[3]	N. a.[7]
(ix′)		No[8]	Yes	Yes	?[5]	No[8]	No

the aggregated uncertainty. We introduced a new measure for aggregated uncertainty, the sum of weighted entropy and Kullback-Leibler divergence, where the weights are second-order probabilities.

Furthermore, we showed that such a measure is equivalent to the entropy of the mean first-order probability distribution. In the paradigm of two levels of uncertainty, the entropy of the mean could be viewed as the amount of uncertainty that can be distributed on the first and second-order levels. From such a perspective it is impossible to unequivocally express unqualified ignorance, i.e. one must declare how the uncertainty is distributed. And unless some reasonable principle of uninformative distribution of uncertainty is formulated we cannot say that first-order uncertainty is more or less uncertain than second-order uncertainty.

We compared the measures by a set of properties that has previously been utilized by [16] and [21]. Our separation of measures into first, second-order and aggregated is not very well reflected in the properties that are held, there is no apparent pattern distinguishing the groups of measures. The relation of measures and properties could be seen as a degree of quality for the measures, the more properties held the better. On the other hand it could be discussed how appropriate the properties are for first and second-order measures of uncertainty.

For example property (viii) seems at odds with the idea that a second-order probability distribution weighs first-order distributions differently. From our perspective it is perfectly reasonable that an uncertainty measure for a second-order distribution is lower than the entropy of any first-order distribution in its support, at least if this distribution has low second-order probability. If a distributions' second-order probability is low then it and its properties have a correspondingly low impact.

If we compare our aggregated measure with the GSU-measure, several differences can be found. Perhaps the

first notably such is that GSU gives ∞ as a result for examples C - E in Table II while our measure are always finite. As previously mentioned, the reason for such a result is due to that a positive second-order probability was assigned to a zero first-order probability. Beside the differences in properties, found in Table III, the GSU-measure also seems to put more emphasis on the first-order level, as is seen from examples A and B.

In example A there is no uncertainty at the second level but a maximum of uncertainty on the first level, while in example B the second-order uncertainty is uniformly distributed around this maximal first-level uncertainty. In this case DI and the GSU-measure gives a aggregated uncertainty that is a bit higher in example B whilst our measure gives the same amount of uncertainty, however, distributed differently among the first and second-order level. As for examples A and B, the pure entropy measure H stands out since it measures the uncertainty of Example A as zero but the uncertainty of B as the maximal $\log_2 3$. This example shows that it could be problematic to use a pure second-order probability entropy measure unless you do believe that belief in three different but close points should reflect much higher uncertainty than belief in a single point. H is also unique among the measures studied here in that it is bounded by the number of first-order distributions in the support of the second-order distributions, giving much higher uncertainty values for Examples C, D and E than the others except GSU.

We have introduced a new dimension of uncertainty but many questions remain. What should be required of measures for first- and second-order uncertainty of second-order distributions? Is it possible to express a larger amount of uncertainty with higher-order distributions? And how is uncertainty on higher levels to be measured?

The assumption of discrete distributions causes many questions. See for instance the discussion above about the measure H, does a finer granularity imply greater uncertainty?

The points in the real-valued probability simplex that have no support in our discrete examples, should they be viewed as non-existent or simply as (second-order) impossible? Under what circumstances can discrete second-order distributions be seen as corresponding to reality, or when can they be justified as approximations or simplifications? And what consequences do such considerations have on the properties that are desired for uncertainty measures? And how could uncertainty measures for continuous second-order distributions be designed and interpreted? These and similar questions are to be addressed in future research.

REFERENCES

[1] R. F. Nau, "Uncertainty aversion with second-order utilities and probabilities," *Management Science*, vol. 52, no. 1, pp. 136–145, 2006.

[2] 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*, 2003, pp. 547–561.

[3] L. Ekenberg and J. Thorbiörnson, "Second-order decision analysis," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 9, No 1*, vol. 9, no. 1, pp. 13–38, 2001.

[4] P. Gärdenfors and N.-E. Sahlin, "Decision, probability and utility: Selected readings." in *Decision, Probability and Utility: Selected Readings*. Cambridge University Press, 1988, ch. 16, Unreliable probabilities, risk taking, and decision making, pp. 313–334.

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

[6] L. A. Zadeh, "Fuzzy probabilities," *Information Processing and Management, 20*, pp. 363–372, 1984.

[7] P. Smets, "Varieties of ignorance and the need for well-founded theories," *Inf. Sci.*, pp. 135–144, 1991.

[8] E. T. Jaynes, "Monkeys, kangaroos, and n," *Maximum Entropy and Bayesian methods in Applied Statistics: proceedings of the 4th Maximum Entropy Workshop*, pp. 26–58, 1984.

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

[10] G. J. Klir, "A principle of uncertainty and information invariance," *International Journal Of General System*, vol. 17, no. 2-3, pp. 249–275, 1990.

[11] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *International Journal of General System*, vol. 9, no. 4, pp. 249–260, 1983.

[12] G. J. Klir and R. M. Smith, "Recent developments in generalized information theory," *International journal of fuzzy systems*, vol. 1, no. 1, pp. 1–13, 1999.

[13] A.-L. Jousselme, C. Liu, D. Grenier, and E. Bosse, "Measuring ambiguity in the evidence theory," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 36, no. 5, pp. 890–903, 2006.

[14] M. Smithson, "Freedom: A measure of second-order uncertainty for intervalic probability schemes," in *Proceedings of the Fifth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-89)*. Corvallis, Oregon: AUAI Press, 1989, pp. 327–334.

[15] P. Gärdenfors and N.-E. Sahlin, "Unreliable probabilities, risk taking, and decision making," *Synthese*, vol. 53, no. 3, pp. 361–386, 1982. [Online]. Available: http://dx.doi.org/10.1007/BF00486156

[16] J. C. Mork, "Uncertainty, credal sets and second order probability," *Synthese*, 2011, qP 20120202.

[17] 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*, 2007, pp. 47–56.

[18] S. Kullback and R. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[19] A. Karlsson, R. Johansson, and S. F. Andler, "Characterization and empirical evaluation of bayesian and credal combination operators," *Journal of Advances in Information Fusion*, vol. 6, pp. 150–166, 2011.

[20] G. Pólya, "Sur quelques points de la théorie des probabilités," *Ann. Inst. Poincaré*, vol. 1, pp. 117–161, 1931.

[21] D. Harmanec and G. J. Klir, "Measuring total uncertainty in dempster-shafer theory: A novel approach," *International Journal of General Systems*, vol. 22, no. 4, pp. 405–419, 1994.