Discrete Second-order Probability Distributions that Factor into Marginals

David Sundgren
University of Gävle
dsn@hig.se

Abstract

In realistic decision problems there is more often than not uncertainty in the background information. As for representation of uncertain or imprecise probability values, second-order probability, i.e. probability distributions over probabilities, offers an option. With a subjective view of probability second-order probability would seem to be impractical since it is hard for a person to construct a second-order distribution that reflects his or her beliefs. From the perspective of probability as relative frequency the task of constructing or updating a second-order probability distribution from data is somewhat easier. Here a very simple model for updating lower bounds of probabilities is employed.

But the difficulties in choosing second-order distributions may be further alleviated if structural properties are considered. Either some of the probability values are dependent in some way, e.g. that they are known to be almost equal, or they are not dependent in any other way than what follows from that the values sum to one.

In this work we present the unique family of discrete second-order probability distributions that correspond to the case where dependence is limited. These distributions are shown to have the property that the joint distributions are equal to normalised products of marginal distributions. The distribution family introduced here is a generalisation of a special case of the multivariate Pólya distribution and is shown to be conjugate prior to a compound hypergeometric distribution.

Keywords. Discrete probability, second-order probability, imprecise probability, multivariate Pólya distribution, conjugate prior, compound hypergeometric likelihood.

1 Introduction

In non-trivial decision problems there is often uncertainty about background data. A decision support system or any system that is meant to work with such uncertain data needs a form of representation for uncertain information or else ignore the uncertainty, i.e. allow for false certainty or false precision. Here we are concerned with representation of uncertain or imprecise probability values. Uncertainty and imprecision will be treated the same way, whether a decision maker believes that there is a precise value but is uncertain as regards to what it is, or if imprecision is inherent, the end result is that there is a set of feasible probability values.

Among models for imprecise probability there are interval based approaches, [8, 9, 14, 19, 20, 21], where the probability of an event is represented by two numbers, the lowest and highest possible value. There are also hierarchical models such as those in [11, 10, 23, 6, 4, 2, 22, 18, 17, 5, 12], where each probability value in the interval is weighed. The potential for discrimination that is present in hierarchical models may be utilized to express that some probability values are more reasonable than others. However, this power is difficult to wield since on the one hand local, one-dimensional, changes have global, multi-dimensional, effects that might be hard to grasp, and on the other hand since given some beliefs about imprecise values there appears to be countless sets of weights that are consistent with the beliefs.

1.1 Structural Considerations

The solution might lie in adding structural information, information that is not asked for in traditional models for imprecise probability, but is nonetheless crucially important and not necessarily hard to extract. The importance of structural information in general is argued for in [3]. Here we will focus on one such property, dependency. Dependency is a fun-

damental concept in probability theory. In this paper we work with one particular hierarchical model, second-order probability, where the weights are themselves probabilities. Since second-order probability is a concept that resides fully inside probability theory there is no reason to assume that issues of dependency would be unimportant in that context.

Now the stochastic variables in a second-order probability distribution are probabilities of events in the same outcome space, so the variables are non-negative and sum to one. This fact alone obviously rules out independence. But two first-order probabilities might be dependent beyond the summing to one, it is conceivable that two probability values are almost the same in all situations, or that if one value increases by a certain amount, another value decreases even more. As an example of the first mentioned case we could take the probabilities of three mutually exclusive events A, B and C. We have that $\Pr(B \cup C) = 1 - \Pr(A)$, but assume that $\Pr(B) = \Pr(C) = x$. Then all probability vectors $(\Pr(A), \Pr(B), \Pr(C))$ have the form $(1 - x, x, x)$ and B and C have a higher degree of dependency than what is prescribed by $\Pr(A) + \Pr(B) + \Pr(C) = 1$. Then again, cases where there are no such further dependencies are also conceivable and this is the case that we explore in this paper since independence usually is less complicated than dependence.

Below we suggest a notion that is intended to capture such limited dependency, i.e. that the probabilistic constraint of non-negative variables summing to one is the only source of dependency. Further we demonstrate that such limited dependency means that the joint second-order probability distribution factors into its own marginal distributions, almost as a joint distribution of independent variables. The difference is that since the variables are not really independent the joint distribution is equal to the normalised product of marginal distributions, the product is multiplied with a factor not equal to one.

A family of continuous second-order probability distribution with this property is in [16] shown to be a shifted or contracted variant of the Dirichlet distribution. In fact, the parameters of that version of the Dirichlet distributions are locked to $1/(n-1)$, where n is the number of possible outcomes, instead a new set of parameters a_i, $i = 1, \ldots, n$, are introduced. The a_i are lower bounds of the first-order probability variables. In other words, the lower bounds determine the distribution. The topic of this paper is the discrete counterpart of the contracted Dirichlet distribution that factors into marginals.

Among the reasons for looking at discrete second-order probability distributions as opposed to the shifted Dirichlet distribution are determination and updating of lower bounds. A lower bound for a probability in a continuous second-order distribution can usually not be the result of an observation, but after seeing one black tulip among 20 in a flower shop I know that the probability of a random tulip in the shop being black is at least 1/20. There could also be computational advantages to discrete distributions and in practice the limited resolution of a discrete distribution might be sufficient.

Since an important advantage of discrete second-order distributions as opposed to their continuous counterparts is that they fit nicely into a simple model for updating we consider the conditions under which the distribution considered here are conjugate. Conjugacy is of interest here since it would be important to know whether the structural properties represented by a family of distributions such as that shown here can remain after updating.

The main result of this paper is then twofold; the unique family of discrete second-order probability distributions that factor into marginals and the compound hypergeometric likelihood that is needed for these distributions to be conjugate.

2 Limited Dependency

We assume that all first-order probability values can be written as a ratio k_i/N, where $k_i \geq 0$ and $\sum_{i=1}^n k_i = N$. For simplicity we will use the nominators k_i as variables, the denominator N would always be the same. We want to capture and formalise the notion that $\sum_{i=1}^n k_i = N, k_i \geq 0$ is the only source of dependency among the variables k_i. When this is the case, the value of a variable would depend on other variables but dependency would only be a function of the sum of variables. For instance, considering the value of k_1 it is important what value the sum of say, k_3 and k_6 holds, but it is irrelevant if k_3 increases and k_6 decreases as long as the sum $k_3 + k_6$ stays the same.

Let $X \not\ni k_i$ be a subset of the set $\{k_1, k_2, \ldots, k_n\}$ of random variables ($\sum_{i=1}^n k_i = N$). By definition of conditional probability $p_i(k_i|X) = \frac{p_i(k_i \cup X)}{p(X)}$. The p:s are probability mass functions, indexed where needed to indicate marginal functions. If we wish k_i:s dependency of X to be limited to a function of the sum of variables we should be able to describe $p_i(k_i|X)$ as

$$p_i(k_i|X) = p_i(k_i) \frac{f(k_i + \sum_{k_i \in X} k_j, |X| + 1)}{f(\sum_{k_i \in X} k_j, |X|)} . \quad (1)$$

In the functions f we need not only sums of variables but also the number of variables in the sum; the value

of a sum of many variables have more information than a sum of few variables even if the sums are equal.

Since

$$p(k_{\pi(1)}, k_{\pi(2)}, \ldots, k_{\pi(n)}) =$$
$$p_{\pi(1)}(k_{\pi(1)})p_{\pi(2)}(k_{\pi(2)}|k_{\pi(1)})$$
$$p_{\pi(3)}(k_{\pi(3)}|k_{\pi(1)}, k_{\pi(2)}) \qquad (2)$$
$$\vdots$$
$$p_{\pi(n)}(k_{\pi(n)}|k_{\pi(1)}, k_{\pi(2)}, \ldots, k_{\pi(n-1)})$$

for any permutation π, if

$$p_{\pi(i)}(k_i|X) = p_{\pi(i)}(k_i)\frac{f(k_i + \sum_{k_i \in X} k_j, |X| + 1)}{f(\sum_{k_i \in X} k_j, |X|)}$$

as in Equation (1) we have that

$$p(k_1, \ldots, k_n) =$$
$$p_{\pi(1)}(k_{\pi(1)})p_{\pi(2)}(k_{\pi(2)})\frac{f(k_{\pi(1)} + k_{\pi(2)}, 2)}{f(k_{\pi(1)}, 1)}$$
$$p_{\pi(3)}(k_{\pi(3)})\frac{f(k_{\pi(1)} + k_{\pi(2)} + k_{\pi(3)}, 3)}{f(k_{\pi(1)} + k_{\pi(2)}, 2)}$$
$$\vdots \qquad (3)$$
$$p_{\pi(n)}(k_{\pi(n)})\frac{f(k_{\pi(1)} + \ldots + k_{\pi(n)}, n)}{f(k_{\pi(1)} + \ldots + k_{\pi(n-1)}, n-1)} =$$
$$\prod_{i=1}^n p_i(k_i)\frac{f(k_{\pi(1)} + \ldots + k_{\pi(n)}, n)}{f(k_{\pi(1)}, 1)}$$

The numerator $f(k_{\pi(1)} + \ldots + k_{\pi(n)}, n)$ is obviously constant since $\sum_{i=1}^n k_i$ is constant equal to N. But the denominator is apparently dependent on the permutation π: if $f(k_i, 1)$ is not constant it is not possible to express the joint probability distribution $p(k_1, \ldots, k_n)$ in this way. On the other hand, if $f(k_i, 1)$ is constant $p(k_1, \ldots, k_n)$ equals the product of marginal distributions multiplied with a constant. That is, if the type of limited dependency described by Equation (1) is achievable the joint probability distribution must factor into marginals.

3 Factoring into Marginals

We have seen that dependence limited to the sum of random variables means that the joint probability density function is proportional to the product of marginal distributions. In the case of discrete second-order probability distributions the limitation is that random variables $k_i, 1 \le i \le n$ are such that $k_i \ge 0$ and $\sum_{i=1}^n k_i = N$. Note that the k_i/N are probabilities, not the k_i. We could have the rational numbers

k_i/N as random variables, but presentation is simplified by dropping the denominator.

Before delving into the calculations, some words about the z transform might be in place. Below we solve the problem at hand by using the convolution property that $\mathcal{Z}\{p_1(k) * p_2(k)\} = \mathcal{Z}p_1(k)\mathcal{Z}p_2(k)$ so that the integrals involved in computing marginal distributions can be computed by eliminating products in a system of equations of products. That we can use convolutions is due to the variables having a fixed sum. The z transform most used below is that of $\frac{\Gamma(k-x+y)}{(k-x)!\Gamma(y)}$ which is $\frac{1}{(1-\frac{1}{z})^y z^x}$. In turn, the Gamma function $\Gamma(x)$ is defined as $\int_0^\infty t^{x-1}e^{-t}\,dt$ for complex numbers with positive real parts. For integers it is just the sg hifted factorial, $\Gamma(n) = (n-1)!$. For more on the z transform, see [7] and on the Gamma function, see e.g. [1]

Dependence limited to the sum of k_i being constant equal to N means that

$$p(k_1, k_2, \ldots, k_n) = \frac{1}{K}\prod_{i=1}^n p_i(k_i),$$

where p_i is the marginal distribution corresponding to variable k_i. Please observe that $\sum_{i=1}^n k_i = N$ throughout the paper.

Then the marginal distribution $p_i(k_i)$ equals

$$\frac{1}{K}p_i(k_i) \underset{j \ne i}{\text{\Large$*$}} p_j(N - k_i), \qquad (4)$$

where $*_{j \ne i}$ is the $n-1$-fold repeated convolution $p_1 * p_2 * \cdots * p_{i-1} * p_{i+1} * \cdots * p_n$ and $K = *_{i=1}^n p_i(N)$.

In the transform domain,

$$\prod_{j \ne i} \mathcal{Z}\{p_j(k_j)\} = \mathcal{Z}\{KH(c_i - k_i)\} \qquad (5)$$

for all $i, i = 1, \ldots, n$, where H is the Heaviside function and the support of p_i ends at $k_i = c_i$. Cancelling in these n equations in the z domain implies that except for different shifts all marginals p_i are equal.

Since the z transform of a constant K is $\frac{Kz}{z-1}$, if $p_i(k_i)$ is any shifted function $q_i(k_i - a_i)$,

$$\mathcal{Z}\{p_i(k_i)\} = \left(\frac{Kz}{z-1}\right)^{\frac{1}{n-1}}\frac{1}{z^{a_i}} \qquad (6)$$

due to the shift property $\mathcal{Z}\{x(n-k)\} = \mathcal{Z}\{x(n)\}z^{-k}$ and

$$\prod_{j \ne i} \mathcal{Z}\{p_j(k_j)\} = \frac{Kz}{z-1}\frac{1}{z^{\sum_{j \ne i} a_j}}, \qquad (7)$$

hence

$$\ast_{j\neq i}p_i(k_i) = \mathcal{Z}^{-1}\left\{\frac{Kz}{z-1}\frac{1}{z^{\sum_{j\neq i}a_j}}\right\}(k_i) =$$

$$KH\left(k_i - \sum_{j\neq i}a_j\right),\qquad(8)$$

giving

$$\ast_{j\neq i}p_i(N-k_i) = KH\left(N-k_i-\sum_{j\neq i}a_j\right)\qquad(9)$$

which equals $KH\left(c_i - k_i\right)$ if $c_i = N - \sum_{j\neq i}a_j$, i.e. the upper limit of the support of p_i is $N - \sum_{j\neq i}a_j$, where a_j is the lower limit of the support of marginal distribution p_j.

So

$$p_i(k_i) = \mathcal{Z}^{-1}\left\{\left(\frac{Kz}{z-1}\right)^{\frac{1}{n-1}}\frac{1}{z^{a_i}}\right\}(k_i) =$$

$$\frac{K^{\frac{1}{n-1}}\Gamma\left(k_i - a_i + \frac{1}{n-1}\right)}{(k_i - a_i)!\Gamma\left(\frac{1}{n-1}\right)}.\qquad(10)$$

And

$$K =$$

$$\ast_{i=1}^n p_i(N) = \mathcal{Z}^{-1}\left\{\prod_{i=1}^n \mathcal{Z}\{p_i(k_i)\}\right\}(N) =$$

$$\mathcal{Z}^{-1}\left\{\prod_{i=1}^n\left(\frac{Kz}{z-1}\right)^{\frac{1}{n-1}}\frac{1}{z^{a_i}}\right\}(N) =$$

$$K^{\frac{n}{n-1}}\mathcal{Z}^{-1}\left\{\left(\frac{z}{z-1}\right)^{\frac{n}{n-1}}\frac{1}{z^{\sum_{i=1}^n a_i}}\right\}(N) =\qquad(11)$$

$$K^{\frac{n}{n-1}}\mathcal{Z}^{-1}\left\{\left(\frac{z}{z-1}\right)^{\frac{n}{n-1}}\right\}(N - \sum_{i=1}^n a_i) =$$

$$K^{\frac{n}{n-1}}\frac{(N-\sum_{i=1}^n a_i)!\Gamma\left(\frac{1}{n-1}\right)}{(n-1)\Gamma\left(N+1-\sum_{i=1}^n a_i + \frac{1}{n-1}\right)}$$

That is,

$$K = \left(\frac{(N-\sum_{i=1}^n a_i)!\Gamma\left(\frac{1}{n-1}\right)}{(n-1)\Gamma\left(N+1-\sum_{i=1}^n a_i + \frac{1}{n-1}\right)}\right)^{n-1}\qquad(12)$$

and the marginal distributions are

$$p_i(k_i) =$$

$$\frac{(N-\sum_{j=1}^n a_j)!\Gamma\left(k_i - a_i + \frac{1}{n-1}\right)}{(n-1)\Gamma\left(N+1-\sum_{j=1}^n a_j + \frac{1}{n-1}\right)(k_i - a_i)!},\qquad(13)$$

$i = 1,\ldots,n$

The joint distribution is

$$p(k_1,\ldots,k_n) =$$

$$\frac{(N-\sum_{i=1}^n a_i)!\prod_{i=1}^n \frac{\Gamma(k_i - a_i + \frac{1}{n-1})}{(k_i - a_i)!}}{(n-1)\Gamma\left(\frac{1}{n-1}\right)^{n-1}\Gamma\left(N+1-\sum_{i=1}^n a_i + \frac{1}{n-1}\right)}\qquad(14)$$

Going back to Section 2 we have now seen that the form of limited dependency that implies factoring into marginals is possible to realise, in fact by considering the multivariate marginal distributions it can be shown that the functions in Equation (3) have the desired properties, that is $f(k_1,\ldots,k_n,n) = 1/K$ and $f(k_i, 1)$ is constant equal to one. The corresponding reasoning could also justify the constraint of factoring into marginals for the contracted Dirichlet distribution of [16].

3.1 Basic Properties

Since $\Gamma(k+x)/k!$ approaches k^{x-1} as k grows when $x << k$, the discrete distribution described above becomes, appropriately normalised, equal to the shifted Dirichlet distribution of [16] when N tends to infinity. In this, k_i/N and a_i/N of the discrete distribution corresponds to the real-valued first-order probability x_i and a_i in the continuous distribution.

Just as the continuous distribution in [16] is a generalization of a Dirichlet distribution with parameters $1/(n-1)$, the discrete probability distribution considered here is, when the parameters $a_i = 0$, a multivariate Pólya distribution [13] with parameters $1/(n-1)$.

The mean of a marginal probability density function $p_i(k_i)$ of the type described here is

$$a_i + \frac{N-\sum_{i=1}^n a_i}{n},\qquad(15)$$

c.f. the mean $a_i + \frac{1-\sum_{i=1}^n a_i}{n}$ of the shifted Dirichlet distribution.

The variance is

$$\frac{(n-1)^2(N-\sum_{i=1}^n a_i)^2}{n^2(2n-1)} + \frac{(n-1)(N-\sum_{i=1}^n a_i)}{n(2n-1)}\qquad(16)$$

which approaches N^2 times the variance of the shifted Dirichlet distribution with lower bounds a_i/N.

The multivariate Pólya distribution is obtained by drawing the underlying probabilities p_i from a Dirichlet distribution and integrating out $\mathbf{p} = (p_1,\ldots,p_n)$ from the multinomial distribution. In the same way, if

we compound the Dirichlet distribution with parameters $1/(n-1)$ with the shifted multinomial distribution

$$\frac{(N - \sum_{i=1}^n a_i)! \prod_{i=1}^n p_i^{k_i - a_i}}{\prod_{i=1}^n (k_i - a_i)!} \tag{17}$$

that is used in [15], we have

$$\int_{\mathbf{P}} \frac{1}{(n-1)^n \Gamma(n/(n-1))^{n-1} \prod_{i=1}^n p_i^{\frac{n-2}{n-1}}}$$
$$\frac{(N - \sum_{i=1}^n a_i)! \prod_{i=1}^n p_i^{k_i - a_i}}{\prod_{i=1}^n (k_i - a_i)!} d\mathbf{p} = \tag{18}$$
$$\frac{(N - \sum_{i=1}^n a_i)!}{(n-1)\Gamma(1/(n-1))^{n-1}\Gamma(N + 1 + 1/(n-1))}$$
$$\prod_{i=1}^n \frac{\Gamma(k_i - a_i + 1/(n-1))}{(k_i - a_i)!}$$

That is, the joint discrete distribution that factors into marginals.

4 Example

Let $n = 4, N = 8$ and $a_1 = 0, a_2 = 1, a_3 = 3, a_4 = 0$. Then

$$p(k_1, k_2, k_3) =$$
$$\frac{4!\Gamma(k_1 + 1/3)\Gamma(k_2 - 1 + 1/3)}{3\Gamma(1/3)^3\Gamma(5 + 1/3)k_1!(k_2 - 1)!}$$
$$\frac{\Gamma(k_3 - 3 + 1/3)\Gamma(8 - k_1 - k_2 - k_3 + 1/3)}{(k_3 - 3)!(8 - k_1 - k_2 - k_3)!}$$

k_1	0	1	2	3	4
	0.534	0.178	0.119	0.0923	00769
k_2	1	2	3	4	5
	0.534	0.178	0.119	0.0923	0.0769
k_3	3	4	5	6	7
	0.534	0.178	0.119	0.0923	0.0769

Table 1: Marginal probability density values for $p_1(k_1), k_1 = 0, \ldots, 4, p_2(k_2), k_2 = 1, \ldots, 5$ and $p_3(k_3), k_3 = 3, \ldots, 7$.

and the marginal distributions are

$$p_1(k_1) = \sum_{k_2=1}^{5-k_1} \sum_{k_3=3}^{8-k_1-k_2} p(k_1, k_2, k_3) =$$
$$\frac{4!\Gamma(k_1 + 1/3)}{3\Gamma(5 + 1/3)k_1!}, \tag{19}$$

$$p_2(k_2) = \sum_{k_1=0}^{5-k_2} \sum_{k_3=3}^{8-k_1-k_2} p(k_1, k_2, k_3) =$$
$$\frac{4!\Gamma(k_2 - 1 + 1/3)}{3\Gamma(5 + 1/3)(k_2 - 1)!}, \tag{20}$$

$$p_3(k_3) = \sum_{k_1=0}^{7-k_3} \sum_{k_2=1}^{8-k_1-k_3} p(k_1, k_2, k_3) =$$
$$\frac{4!\Gamma(k_3 - 3 + 1/3)}{3\Gamma(5 + 1/3)(k_3 - 3)!}, \tag{21}$$

$$p_4(k_4) = p_4(8 - k_1 - k_2 - k_3) =$$
$$\sum_{k_1=0}^{4-k_4} \sum_{k_2=1}^{5-k_1-k_4} p(k_1, k_2, 8 - k_1 - k_2 - k_4) =$$
$$\frac{4!\Gamma(k_4 + 1/3)}{3\Gamma(5 + 1/3)k_4!} = \frac{4!\Gamma(8 - k_1 - k_2 - k_3 + 1/3)}{3\Gamma(5 + 1/3)(8 - k_1 - k_2 - k_3)!} \tag{22}$$

The means of p_1, p_2, p_3 and p_4 are $1, 2, 4$ and 1, respectively, corresponding to mean first-order probabilities of $1/8, 1/4, 1/2$ and $1/8$. Since k_1 and k_4 share the same conditions, their respective marginal probability density functions are equal. We see a table with values of the marginal distribution functions in Table 1. The values reveal that the distributions are essentially equal but differently shifted according to their respective lower bounds of support.

5 Updating

One advantage of treating relative frequencies as first-order probabilities is that updating of lower bounds of probabilities may come about in a natural way. See [15], where this is discussed and exemplified with (shifted) multinomial distributions as prior and posterior distributions and a hypergeometric likelihood. In

this paper we are concerned with a shifted version of the multivariate Pólya distribution where the parameters of the Pólya distribution are locked at $1/(n-1)$ but a new vector $\begin{pmatrix} a_1 & a_2 & \ldots & a_n \end{pmatrix}$ of parameters is introduced, where the a_i are lower bounds, i.e. for whatever reason we know that there are at least a_i objects of type i among the total N objects.

As described above in Section 3 this variant of the multivariate Pólya distribution represents a situation where the variables k_i, the number of objects of respective type, are in a sense minimally dependent. This property does not necessarily remain after updating and since the case of further dependencies than those incurred by $\sum_{i=1}^n k_i = N$ remains to be investigated we choose to consider the conditions under which updating must be done for the shifted multinomial Pólya distribution to be a conjugate distribution. First though, the model for updating deserves some explanation.

5.1 The Urn and the Plate

Since the lower bounds a_i are the only parameters it is these values that can be affected by updating. The idea behind the model proposed in [15] is that if I observe a_i objects of type i I know with absolute certainty that there were at least a_i such objects to begin with. In terms of the ubiquitous urn, we have N balls with n different colours in an urn and the question is as usual how many balls there are of each colour in the urn. Updating consists of picking a handful $(\sum_{i=1}^n a_i)$ of balls from the urn and observing that a_i of them have colour i.

Then we know that there were at least a_i balls with colour i in the urn to begin with. But in terms of probabilities and relative frequencies we are only interested in these numbers in relation to the original number N of balls in the urn, e.g. after observing three green balls from an urn with 20 balls I know that the relative frequency of green balls in the urn was at least $3/20$. Thus one might think that replacement is in order so that there remains N balls in the urn. However, if I after replacement pick three green balls again in the next round I have no justification for claiming that there at least six green balls out of 20 since some of the balls might be the same as in the previous updating. One solution could be to mark the already observed balls and ignore them in future updating but then I would not know the results of previous experiments without taking notes. Putting the observed balls on a plate on the side in full sight saves ink and paper and reminds us that observed balls are not simply not replaced in the sense of being discarded. The balls on the plate count but updating is only done by probing the urn.

5.2 Shifted Pólya as Conjugate Prior

First let us observe that since the discrete second-order distributions that are topic of this paper factor into marginals, if prior and posterior are both from this family, the likelihood must also factor into marginals. We look at the one-variable marginal case first for ease of presentation. W.l.o.g. we assume that the prior distribution have parameters $a_i = 0$, i.e. nothing has been observed and apart from a structural assumption of minimal dependency we know nought but N, the total number of objects in the urn, and n, the number of different colours. As described in Section 5.1 above the experiment consists of drawing $\sum_{i=1}^n a_i$ balls from the urn and thus rule out the possibility that the number k_i of balls with colour i would be less than a_i.

The i:th marginal of the prior is Beta-binomial with parameters $\alpha = \frac{1}{n-1}$ and $\beta = 1$. i.e.

$$\binom{N}{k_i} \frac{\mathrm{B}\left(k_i + \frac{1}{n-1}, N - k_i + 1\right)}{\mathrm{B}\left(\frac{1}{n-1}, 1\right)}, \qquad (23)$$

the i:th marginal of the posterior is Beta-binomial with the same parameters $\alpha = \frac{1}{n-1}, \beta = 1$ as in the prior but k_i replaced with $k_i - a_i$ and N substituted for $N - \sum_{j=1}^n a_j$:

$$\frac{\binom{N - \sum_{j=1}^n a_j}{k_i - a_i} \mathrm{B}\left(k_i - a_i + \frac{1}{n-1}, N - \sum_{j\neq i} a_i - k_i + 1\right)}{\mathrm{B}\left(\frac{1}{n-1}, 1\right)}. \qquad (24)$$

The corresponding likelihood is achieved by a weighted hypergeometric distribution

$$\frac{\binom{N-\sum_{j=1}^n a_j}{k_i - a_i}}{\binom{N}{k_i}} p^{-a_i}(1-p)^{a_i - \sum_{j=1}^n a_j}, \qquad (25)$$

where p is drawn from $\mathrm{Beta}\left(k_i + \frac{1}{n-1}, N - k_i + 1\right)$ so that the likelihood is the compound distribution

$$\int_0^1 \frac{\binom{N-\sum_{j=1}^n a_j}{k_i-a_i}}{\binom{N}{k_i}} p^{-a_i}(1-p)^{a_i-\sum_{j=1}^n a_j}$$

$$\frac{p^{k_i-\frac{n-2}{n-1}}(1-p)^{N-k_i}}{\mathrm{B}\left(k_i+\frac{1}{n-1},N-k_i+1\right)}\,dp =$$

$$\frac{\binom{N-\sum_{j=1}^n a_j}{k_i-a_i}}{\binom{N}{k_i}} \qquad (26)$$

$$\frac{\mathrm{B}\left(k_i-a_i+\frac{1}{n-1},N-\sum_{j=1}^n a_j+a_i-k_i+1\right)}{\mathrm{B}\left(k_i+\frac{1}{n-1},N-k_i+1\right)}$$

The multivariate likelihood is the weighted hypergeometric distribution

$$\prod_{i=1}^n \frac{\binom{N-\sum_{j=1}^n a_j}{k_i-a_i}}{\binom{N}{k_i}} p_i^{-a_i}\,, \qquad (27)$$

where **p** is drawn from the Dirichlet distribution with parameters $k_i+\frac{1}{n-1}$. That is,

$$\prod_{i=1}^n \frac{\binom{N-\sum_{j=1}^n a_j}{k_i-a_i}}{\binom{N}{k_i}} \frac{\Gamma\left(\sum_{i=1}^n k_i+\frac{1}{n-1}\right)}{\prod_{i=1}^n \Gamma\left(k_i+\frac{1}{n-1}\right)}$$

$$\int_{\mathbf{P}} \prod_{i=1}^n p_i^{k_i-a_i+\frac{1}{n-1}-1}\,d\mathbf{p} = \qquad (28)$$

$$\Gamma\left(N+\frac{1}{n-1}\right)\left(\left(N-\sum_{j=1}^n a_j\right)!\right)^n$$

$$\prod_{i=1}^n \frac{k_i!(N-k_i)!\Gamma\left(k_i-a_i+\frac{1}{n-1}\right)}{N!\left(N-\sum_{j=1}^n a_j-k_i+a_i\right)!(k_i-a_i)!\Gamma\left(k_i+\frac{1}{n-1}\right)}$$

Admittedly this likelihood function appears rather exotic, particularly in the factors $p_i^{-a_i}$ which mean that it is more likely to draw a larger number of balls of a certain colour. In contrast, as seen in Section 4 the prior and posterior distributions are such that lower values of the number of objects of type i have higher probability. The full implications of this are yet to be considered but one possible interpretation is that such likelihood functions would rarely be seen in nature as it were. In that case the limited dependency of the original proportions in the urn is fragile and easily disturbed when removing objects.

6 Conclusions

Structural properties such as dependency might be worth considering when choosing a second-order distribution for the purpose of expressing imprecise probabilities. Second-order probability distributions have

probability values as variables, hence independence is impossible. We have however suggested that joint second-order probability distributions that are equal to the normalised products of their own marginal distributions capture the property of a form of minimal dependency. A continuous family of second-order distributions has been found earlier but here a corresponding discrete family is discovered. This family can be described as a generalisation of a special case of the multivariate Pólya distribution where the parameters are fixed but new parameters in the form of lower bounds on the variables are introduced.

The raison d'être of discrete second-order distributions is that they allow for interpreting relative frequencies as first-order probabilities in a natural way. Such a context makes the interpretation of the meaning of second-order probability values easier in that concrete examples in the form of urn models etc. are readily available. An example is updating where a so-called urn-and-plate model gives a simple description of updating of lower bounds. Discrete second-order distributions are also versatile since they apart from relative frequencies also lend themselves to subjective probabilities. That is as long as the subjective probabilities do not involve statements about irrational numbers such as "I am sure that the probability is at least $1/\pi$". Reasonably the lower bound could be given as $8/25$ or some other rational number instead, infinite precision is meaningless in subjective probability judgements.

The family of distributions discussed here represents a form of limited dependency. We have seen that the family being conjugate requires a rather special likelihood function which suggests that the property of limited dependency is sensitive to the removal of objects that occurs in updating of lower bounds in the plate-and-urn model. Full understanding of the meaning of the parameters of the compound likelihood is however a matter for further investigation.

Acknowlededgments

References

[1] R. A. Asker and R. Roy. *NIST handbook of mathematical functions*, chapter Gamma function. Cambridge University Press, 2010.

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

[3] 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*, pages 62–67. AAAI Press, 2007.

[4] 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.

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

[6] 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.

[7] Eliahu Jury. *Theory and application of the z-transform method*. R.E. Krieger Pub. Co, 1973.

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

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

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

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

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

[13] G. Pólya. Sur quelques points de la théorie des probabilités. *Ann. Inst. Poincaré*, 1:117–161, 1931.

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

[15] D. Sundgren. Expected utility from multinomial second-order probability distributions. *Polibits*, (42):71–75, 2010.

[16] D. Sundgren, L. Ekenberg, and M. Danielson. 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*, pages 405–410, 2009.

[17] 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.

[18] 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.

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

[20] P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2-3):125–148, 2000.

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

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

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