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

David Sundgren
University of Gävle
dsn@hig.se

Love Ekenberg
Stockholm University
and KTH
lovek@dsv.su.se

Mats Danielson
Stockhom University
and KTH
mad@dv.su.se

Abstract

In classic decision theory it is assumed that a decision-maker can assign precise numerical values corresponding to the true value of each consequence, as well as precise numerical probabilities for their occurrences. In attempting to address real-life problems, where uncertainty in the input data prevails, some kind of representation of imprecise information is important. Second-order distributions, probability distributions over probabilities, is one way to achieve such a representation. However, it is hard to intuitively understand statements in a multi-dimensional space and user statements must be provided more locally. But the information-theoretic interplay between joint and marginal distributions may give rise to unwanted effects on the global level. We consider this problem in a setting of second-order probability distributions and find a family of distributions that normalised over the probability simplex equals its own product of marginals. For such distributions, there is no flow of information between the joint distributions and the marginal distributions other than the trivial fact that the variables belong to the probability simplex.

Keywords. Second-order probability distribution, Dirichlet distribution, Beta distribution, Kullback-Leibler divergence, relative entropy, product of marginal distributions.

1 Introduction

In attempting to address real-life decision problems, where uncertainty about data prevails, some kind of representation of imprecise information is important and several have been proposed. In particular, first-order representations, such as sets of probability measures [9], upper and lower probabilities [2], and interval probabilities and utilities of various kinds, see e.g. [15, 16], have been suggested for enabling a better representation of the input sentences for a subsequent decision analysis. To facilitate a better qualification of the various possible functions, higher-order estimates, such as distributions expressing various beliefs, can be introduced over n-dimensional spaces, where each dimension corresponds to possible probabilities of events or utilities of consequences. Such hierarchical model approaches are sometimes better suited for modelling incomplete knowledge and can add important information when handling aggregations of imprecise representations, as is the case in decision trees or probabilistic networks [3]. There are, however, at least two problems herein. Firstly, a normal decision maker cannot have any meaningful intuition regarding a multi-dimensional space and the information must be provided more locally, and secondly, it is hard to obtain global information from such local information. Of particular interest in this context is therefore to investigate the relation between global and local distributions.

We will use second-order probabilities, formally defined below in Definition 4, in short, these are probability distributions on random variables that take values on $[0, 1]$ and sum to 1. The intuition of a second-order probability distribution is that it is a distribution that assigns probabilities to the probabilities of the possible outcomes of an event. So such a distribution will have to be defined on the hyper-surface defined by $\sum_{i=1}^{n} x_i = 1, x_i \geq 0, i = 1, \ldots, n$ or, equivalently on the $n-1$-dimensional simplex where $\sum_{i=1}^{n-1} x_i \leq 1, x_i \geq 0, i = 1, \ldots, n-1$ and x_n is an abbreviation of $1 - \sum_{i=1}^{n-1} x_i$. In this paper we will only consider continuous distributions.

The uniform distribution with support on the simplex where $\sum_{i=1}^{n-1} x_i \leq 1, x_i \geq 0, i = 1, \ldots, n-1$ with constant value the inverse of the volume of the simplex and the Dirichlet distribution are examples of second-order probability distributions.

Such second-order probability distributions is one way of handling uncertainty of probabilities in a decision situation, see e.g. [11], [14] and [5]. Instantly, new difficulties appear; on the one hand it may seem that there are too many distributions to choose from given the available knowledge, on the other hand it is not certain that any set of univariate second-order distributions is consistent with the fact that the variables are themselves probabilities. Even if they are consistent, the marginal distributions and the joint distribution may represent different information, e.g. it is shown in [13] that the uniform joint distributions have marginal distributions that are far from uniform.

1.1 Definitions

Definition 1 *[1] For a k-dimensional random vector $(X_1, \ldots, X_k)$ the (joint)* distribution *μ is defined by*

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

where $\mathcal{R}^k$ 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 2 *[1] 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 nonnegative Borel function on R^k and*

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

Definition 3 *[1] 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 $A \in \mathcal{R}$.*

Definition 4 *A* second-order probability distribution *is a distribution μ with support on a set $\mathcal{P} = \{(x_1, \ldots, x_k) : 0 \leq a_i \leq x_i \leq b_i, i = 1, \ldots, k, \sum_{i=1}^k \leq 1\}$.*

1.2 The Problem

Below we will only consider densities and for simplicity abuse terminology as to identify distributions with their probability density functions.

For most decision makers it would be easiest to consider univariate distributions since it is harder to think in several dimensions [4]. In general, though, the marginal distributions together contain more information than the corresponding multivariate distribution. The random variables are the probabilities of the possible outcomes of an event. If the variables are dependent in other than relating to the same event this information discrepancy between local and global is natural since information would be shared between the local variables. But settings where the opposite holds comes easier to mind, and such cases would be better modelled with random variables that are as independent as possible modulo that they sum to one.

The above reasoning motivates us to consider whether there are joint second-order probability distributions that have the same information content as its univariate marginal distributions. This condition will be seen to be equivalent to the joint probability distribution function being equal to the product of its own univariate marginal distributions multiplied with a normalising constant that comes from us working in the probability simplex rather than in the unit cube. That is, the information-theoretic constraint of not losing information when taking marginals coincides with the practical concern of being able to construct a joint probability density from given marginals in the simplest possible way. In terms of copulas (see e.g. [10] or [12]), the condition is that the copula is the product copula multiplied by some constant.

We show that the condition of a joint probability distribution function being equal to the product of its own univariate marginal distributions multiplied with a normalising constant is met by a family of distributions that have the same shape as the Dirichlet distribution. The first-order probability variables can be given arbitrary bounds only from below. When the lower bounds are zero, we have a special case of the Dirichlet distribution where all parameters are equal. With general lower bounds $x_i \geq a_i$, the support of x_i is the interval $[a_i, 1 + a_i - \sum_{i=1}^n a_i]$, the joint Dirichlet distribution and the corresponding marginal Beta distribution are shifted and re-scaled accordingly.

2 Minimal Kullback-Leibler Divergence

To capture the notion that no information other than that of being on the probability simplex is either lost or gained when going between a joint probability distribution and its marginals, we use the *Kullback-Leibler divergence* or *relative entropy* [8], see also [6, 7].

Definition 5 *If P and Q are probability measures over a set X and if μ is a measure such that $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ exist the* Kullback-Leibler divergence *from P to Q is*

$$D_{\mathrm{KL}} \left(P \| Q \right) = \int_X p \log \frac{p}{q} \, d\mu \, .$$

Since we want, as far as possible given that we are on the probability simplex $\mathcal{P}$, that the joint distribution $f(x_1, \ldots, x_n)$ contains the same information as the product of marginals $\prod_{i=1}^{n} f_i(x_i)$, we want the Kullback-Leibler divergence $D_{\mathrm{KL}}(f \| \prod_{i=1}^{n} f_i)$, also known as the *total correlation* [17] of $X_1, \ldots, X_n$ to be minimal. *Gibbs' inequality* states that $D_{\mathrm{KL}}(P \| Q) \geq 0$ with equality only if $P = Q$. Since the probability simplex $\mathcal{P}$ is measurable we can calculate $D_{\mathrm{KL}}(f \| \prod_{i=1}^{n} f_i)$ as a Lebesgue integral.

But restricting the support of $\prod_{i=1}^{n} f_i$ to the probability simplex $\mathcal{P}$ means that $\prod_{i=1}^{n} f_i$ must be normalised in order to be a distribution, i.e. we must find a real number K such that $\int_{\mathcal{P}} \prod_{i=1}^{n} f_i(x_i)/K \, d\mathbf{x} = 1$. So minimising the Kullback-Leibler divergence of $\prod_{i=1}^{n} f_i$ from the joint probability distribution entails finding f such that

$$f(x_1, \ldots x_n) = \frac{1}{K} \prod_{i=1}^{n} f_i(x_i) \, ,$$

where $f_i(x_i)$ is the marginal distribution of $f(\mathbf{x})$ with respect to x_i and $K = \int_{\mathcal{P}} \prod_{i=1}^{n} f_i(x_i) \, d\mathbf{x}$. Let us say that such distributions *factors into marginals.*

3 Characterisation of Distributions that Factors into Marginals

Theorem 1 *A probability distribution $f(\mathbf{x})$ factors into marginals if and only if its marginal distributions are*

$$f_i(x_i) = \frac{1}{(n-1)\left(1 - \sum_{j=1}^{n} a_j\right)^{\frac{1}{n-1}} (x_i - a_i)^{\frac{n-2}{n-1}}}$$

with support $[a_i, 1 - \sum_{j \neq i} a_j]$, where $\sum_{j=1}^{n} a_j < 1$.

Corollary 1 *A joint probability distribution function $f(\mathbf{x})$ on the probability simplex $\mathcal{P}$ factors into marginals if and only if*

$$f(x_1, \ldots, x_{n-1}) = \frac{(1 - \sum_{i=1}^{n} a_i) \prod_{i=1}^{n} f_i(x_i)}{\Gamma^{1-n}\left(\frac{n}{n-1}\right)} \, ,$$

where $x_n = 1 - \sum_{i=1}^{n-1} x_i$, $f_i(x_i), i = 1, \ldots, n$ are the marginal distributions of f and $f_i(x_i) = 0$ for $x_i \geq a_i, x_i \leq 1 - \sum_{j \neq i} a_j$.

When the first-order probability variables x_i are minimally restricted, i.e. $a_i = 0$, $\mathbf{x} = (x_1, \ldots, x_n)$ are Dirichlet distributed with parameters $\alpha_i = \frac{1}{n-1}$ and the marginal distributions f_i are Beta distributions $f(x; \alpha, \beta)$ with parameters $\alpha = \frac{1}{n-1}$ and $\beta = 1$. The marginal distributions f_i also have the same shape as Pareto distributions, but cut off so that the support has upper bound $1 - \sum_{j \neq i} a_j$ rather than infinity.

We make a quick note on the degenerate case where $\sum_{j=1}^{n} a_j = 1$; then the marginal distributions are Dirac pulses $f_i(x_i) = \delta(x_i - a_i)$, i.e. all belief is concentrated in the points $x_i = a_i$ and the joint probability distribution is $\prod_{i=1}^{n} \delta(x_i - a_i)$.

We proceed with the proof of Theorem 1. The proof is based on the fact that an integral $\int_{\mathcal{P}} \prod_{i=1}^{n} g_i(x_i) \, d\mathbf{x}$ of a product of univariate functions over the probability simplex $\mathcal{P}$ is the repeated convolution $g_1 * g_2 * \cdots * g_n(1)$. E.g. when $n = 3$ we have

$$\int_0^1 \int_0^{1-x_1} g *_1 (x_1) g_2(x_2) g_3(1 - x_1 - x_2) \, dx_2 \, dx_1 = \int_0^1 g_1(x_1)[g_2 * g_3(1 - x_1)] \, dx_1 = g_1 * g_2 * g_3(1) \, .$$

If $f(\mathbf{x})$ factors into marginals the marginal distribution with respect to x_i is

$$\frac{1}{K} f_i(x_i) \mathop{\ast}_{j \neq i} f_j \, (1 - x_i) \, ,$$

where $*_{i \neq j} f_j$ is the $n-1$-fold repeated convolution $f_1 * f_2 * \cdots * f_{i-1} * f_{i+1} * \cdots * f_n$ and K is the n-fold convolution $*_{i=1}^{n} f_i(1)$. Assume that $\{f_i\}_{i=1}^{n}$ are the marginal distributions of a joint distribution that factors into marginals. Then for all $i, i = 1, \ldots, n$,

$$\mathop{\ast}_{j \neq i} f_j(1 - x_i) = KH(c_i - x_i) = KH((1 - x_i) - (1 - c_i)) \, ,$$

where c_i is such that $f_i(x_i) = 0$ when $x_i > c_i$.

Then the distributions f_k must have Laplace transforms F_k such that

$$\prod_{k \neq i} F_k = \frac{K e^{-(1 - c_i)s}}{s}$$

and if f_k is on the form $g_k(x_k - a_k) H(x_k - a_k)$ where $f_k(x_k) = 0$ when $x_k < a_k$, g_k must have Laplace transform $\left(\frac{K}{s}\right)^{\frac{1}{n-1}}$, that is $g_k(x_k) = \frac{K^{\frac{1}{n-1}}}{\Gamma\left(\frac{1}{n-1}\right) x_k^{\frac{n-2}{n-1}}}$ and

$$f_k(x_k) = \frac{K^{\frac{1}{n-1}} H(x_k - a_k)}{\Gamma\left(\frac{1}{n-1}\right) (x_k - a_k)^{\frac{n-2}{n-1}}}$$

since the Laplace transform of t^α is $\frac{\Gamma(1+\alpha)}{s^{1+\alpha}}$, where $\Gamma(1 + \alpha) = \int_0^\infty e^{-x} x^\alpha \, dx$.

Further, since the Laplace transform of $f_k(x_k)$ is $\frac{K^{\frac{1}{n-1}} e^{-s a_k}}{s^{\frac{1}{n-1}}}$,

$$\mathop{\ast}_{j \neq i} f_j(1 - x_i) = \frac{K e^{-s(\sum_{j \neq i} a_j)}}{s} \, ,$$

the upper limit of the support of x_i is $c_i = 1 - \sum_{j \neq i} a_j$ and the n-fold convolution $\mathop{\ast}_{i=1}^{n} f_i(t)$ is the inverse

Laplace transform of $\frac{K^{\frac{n}{n-1}} e^{-s\sum_{i=1}^n a_i}}{s^{\frac{n}{n-1}}}$, i.e. $\mathop{\ast}_{i=1}^{n} f_i(t) = \frac{K^{\frac{n}{n-1}} H(t-\sum_{i=1}^n a_i)(t-\sum_{i=1}^n a_i)^{\frac{1}{n-1}}}{\Gamma\left(\frac{n}{n-1}\right)}$, so

$$K = \mathop{\ast}_{i=1}^{n} f_i(1) = \frac{K^{\frac{n}{n-1}} (1-\sum_{i=1}^n a_i)^{\frac{1}{n-1}}}{\Gamma\left(\frac{n}{n-1}\right)}$$

and $K = \frac{\Gamma^{n-1}\left(\frac{n}{n-1}\right)}{1-\sum_{i=1}^n a_i}$.

But since $\Gamma(z+1) = z\Gamma(z)$, $\Gamma\left(\frac{n}{n-1}\right) = \frac{1}{n-1}\Gamma\left(\frac{1}{n-1}\right)$ and

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

4 Some Properties of Second-Order Distributions that Factors into Marginals

The second-order probability distributions that factors into marginals are, as we have seen above, determined by the n-dimensional vector $(a_1, \ldots, a_n)$, where a_i is the lower bound of the support of the marginal distribution f_i. Thus we can by Corollary 1 define its probability density function with respect to the Lebesgue measure as

$$f(x_1, \ldots, x_{n-1}; a_1, \ldots, a_n) = \frac{(1-\sum_{i=1}^n a_i)\prod_{i=1}^n f_i(x_i)}{\Gamma^{1-n}\left(\frac{n}{n-1}\right)},$$

for $x_1, \ldots, x_{n-1}$ such that $\sum_{i=1}^n x_i \le 1$ and where $\sum_{i=1}^n a_i < 1$.

Likewise the marginal distributions are

$$f_i(x_i, a_1, \ldots, a_n) = \frac{1}{(n-1)(1-\sum_{i=1}^n a_i)^{\frac{1}{n-1}} (x_i - a_i)^{\frac{n-2}{n-1}}}$$

with support $[a_i, 1-\sum_{j\neq i} a_j]$. When $a_i = 0$ for all $i = 1, \ldots, n$, $(x_1, x_2, \ldots, x_{n-1})$ have the Dirichlet distribution $f(x_1, \ldots, x_n; 1/(n-1), \ldots, 1/(n-1))$ and the individual variables x_i have Beta distributions $f(x; 1/(n-1), 1)$.

Regarding the intervals of support, one may choose the lower bounds a_i freely as long as the sum $a_1 + \cdots + a_n$ is less than one (and lower bounds summing to a number greater than one is unreasonable since the variables x_i have a sum less than one). But if we want a joint second-order distribution that factors into marginals, the upper bounds are determined by the lower bounds a_i. A consequence of this is that arbitrary support intervals are not in general possible to reconcile with this type of distributions. If the support intervals of the marginal distributions are $[a_i, b_i]$, we cannot form the joint second-order probability distribution as the normalised product of the marginal distributions unless $b_i = 1 - \sum_{j\neq i} a_i$.

Let us list some properties of the marginal distributions; if a second-order probability distribution f with parameters $a_1, \ldots a_n$ factors into marginals the univariate marginal distributions have

- mean $a_i + \frac{1-\sum_{j=1}^n a_j}{n}$,
- median $a_i + \left(\frac{1-\sum_{i=1}^n a_i}{2}\right)^{n-1}$ and
- variance $\frac{(n-1)^2}{n^2(2n-1)}\left(1-\sum_{j=1}^n a_j\right)^2$.

4.1 Multivariate Marginal Distributions

We may generalise the argument in the proof of Theorem 1 to achieve the multivariate marginal distribution of $x_1, \ldots x_k, k < n$ as

$$\frac{1}{K}\prod_{i=1}^k f_i(x_i) \mathop{\ast}_{i=k+1}^{n} f_i\left(1-\sum_{i=1}^k x_i\right).$$

Since the Laplace transform of $f_i(x_i)H(x_i - a_i)$ is $\frac{K^{\frac{1}{n-1}} e^{-sa_i}}{s^{\frac{1}{n-1}}}$ with $K = \frac{\Gamma^{n-1}\left(\frac{n}{n-1}\right)}{1-\sum_{i=1}^n a_i}$ we have the following Corollary.

Corollary 2 *If $f(x_1, \ldots, x_{n-1})$ is a second-order probability distribution that factors into marginals, the multivariate marginal distribution $f(x_1, x_2, \ldots, x_k)$ is*

$$\frac{\prod_{i=1}^k f_i(x_i)(1-\sum_{i=1}^n a_i)^{\frac{k-1}{n-1}}}{\Gamma\left(\frac{n-k}{n-1}\right)\Gamma^{k-1}\left(\frac{n}{n-1}\right)\left(1-\sum_{i=1}^k x_i - \sum_{i=k+1}^n a_i\right)^{\frac{k-1}{n-1}}}.$$

Corollary 2 in turn gives us a result on conditional distributions.

Corollary 3 *The conditional distribution of x_k given $x_1, x_2, \ldots, x_{k-1}$ is*

$$C\frac{f_k(x_k)\left(1-\sum_{i=1}^{k-1} x_i - \sum_{i=k}^n a_i\right)^{\frac{k-2}{n-1}}}{\left(1-\sum_{i=1}^k x_i - \sum_{i=k+1}^n a_i\right)^{\frac{k-1}{n-1}}},$$

where

$$C = \frac{\Gamma\left(\frac{n-k+1}{n-1}\right)(1-\sum_{i=1}^n a_i)^{\frac{1}{n-1}}}{\Gamma\left(\frac{n}{n-1}\right)\Gamma\left(\frac{n-k}{n-1}\right)}$$

if $x_i, i = 1, \ldots n-1$, are distributed by a second-order probability distribution that factors into marginals.

5 Examples

Example 1 *With $n = 3$, let us take $a_1 = 1/3, a_2 = 1/5$ and $a_3 = 1/8$. Then $1 - \sum_{i=1}^{3} a_i = \frac{120-40-24-15}{120} = \frac{41}{120}$.*

$$f_1(x_1) = \frac{1}{2\sqrt{41/120(x_1 - 1/3)}},$$

$$f_2(x_2) = \frac{1}{2\sqrt{41/120(x_2 - 1/5)}}$$

and

$$f_3(x_3) = \frac{1}{2\sqrt{41/120(x_3 - 1/8)}},$$

with support $[1/3, 27/40], [1/5, 13/24]$ and $[1/8, 7/15]$ and mean $\frac{161}{360}, \frac{113}{360}$ and $\frac{43}{180}$, respectively.

The joint distribution $f(x_1, x_2)$ is

$$\frac{41 f_1(x_1) f_2(x_2) f_3(1 - x_1 - x_2)}{120\Gamma^2(3/2)} =$$

$$\frac{\sqrt{120/41}}{\Gamma^2(3/2)\sqrt{(x_1 - 1/3)(x_2 - 1/5)(7/8 - x_1 - x_2)}},$$

see Figure 1 for a plot.

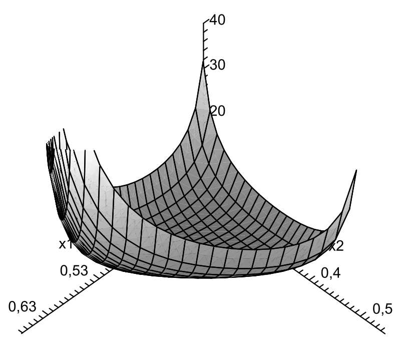

Figure 1: The joint probability density function of Example 1

Given that we in Example 1 wanted to represent knowledge about lower bounds on probabilities, the joint and marginal distribution seem rather rich in information and far from uniform. But we do not wish to minimise entropy in either the joint or the marginal distributions, instead the goal is to balance the entropy of the joint distributions and the marginals. The local effects of maximising entropy globally have partly been studied in [13]. To further study this effect and the converse global effect of local entropy maximisation is a topic for future research.

Example 2 *Let $n = 5$ and $a_i = 1/10$. Then f_1, f_2, f_3, f_4, f_5 where $f_i(x) = f(x) = \frac{1}{4(1/2)^{1/4}(x-1/10)^{1/4}} = \frac{1}{2^{7/4}(x-1/10)^{3/4}}$ are the marginal distributions of a second-order distribution that factors into its own marginals, in Figure 2 we see a plot of the marginal distributions.*

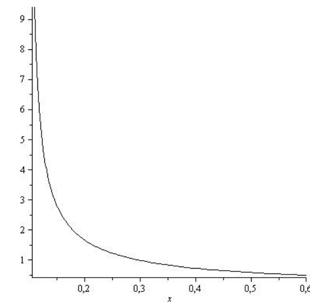

Figure 2: The marginal probability density functions $f_i(x) = 2^{-7/4}(x - 1/10)^{-3/4}$ of Example 2 for $1/10 \leq x \leq 3/5$.

The support of $f_i(x_i)$ is $[1/10, 3/5]$. The means are $\mu_i = 1/5$ and

$$f(x_1, x_2, x_3, x_4) = \frac{(1 - 1/2)\prod_{i=1}^{5} f_i(x_i)}{\Gamma^4(5/4)}.$$

The three variable marginal distribution $f(x_1, x_2, x_3)$ is

$$\frac{f(x_1)f(x_2)f(x_3)}{\Gamma(1/2)\Gamma^2(5/4)\sqrt{2}\sqrt{4/5 - x_1 - x_2 - x_3}},$$

and the conditional distribution $f(x_4|x_1, x_2, x_3)$ is

$$\frac{f(x_4)\Gamma(1/2)\sqrt{4/5 - x_1 - x_2 - x_3}}{2^4\Gamma(5/4)\Gamma(1/4)(9/10 - x_1 - x_2 - x_3 - x_4)^{3/4}}$$

E.g. The conditional distribution $f(x_4|x_1 = 1/10, x_2 = 1/5, x_3 = 2/5)$ is shown in Figure 3.

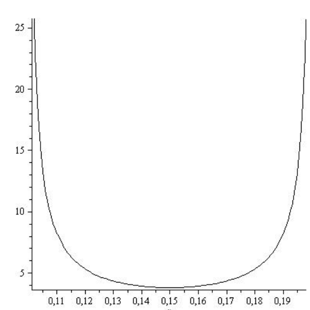

Figure 3: The conditional density of x_4 given $x_1 = 1/10, x_2 = 1/5, x_3 = 2/5$ in Example 2

The two variable marginal distribution $f(x_2, x_4)$ *is*

$$\frac{f(x_2)f(x_4)}{2^{1/4}\Gamma(3/4)\Gamma(5/4)(7/10 - x_2 - x_4)^{1/4}},$$

and the conditional distribution $f(x_1|x_2, x_4)$ *is*

$$\frac{f_1(x_1)\Gamma(3/4)(7/10 - x_2 - x_4)^{1/4}}{2^{1/4}\Gamma(5/4)\Gamma(1/2)(3/5 - x_1 - x_2 - x_4)^{1/2}}$$

The variance of f_i *is*

$$\frac{4^2}{5^2 9}(1/2)^2 = \frac{4}{225}.$$

6 Conclusion

We have found a characterisation of the second-order probability distributions that can be expressed as a normalised product of its own marginal distributions. For such distributions there is a direct path from local to global information. From an information-theoretical standpoint, such probability distributions are unique in that given that the variables are probabilities, no information is either lost or gained when going between the joint distribution and the univariate marginal distributions.

The family of distributions with the properties mentioned above can be said to be a generalisation of a special case of the Dirichlet distribution. When all lower bounds on the first-order probabilities are zero, we get the Dirichlet distribution will all parameters equal to $1/(n-1)$, where n is the number of possible outcomes whose probabilities are the variables. In this case, of course, the marginal distributions are Beta distributions. But in general, with first-order probability variables bounded from below by positive numbers, we have shifted and re-scaled versions of the Dirichlet and Beta distributions, respectively. It is a matter for future research to investigate to which degree properties of Dirichlet and Beta distributions carry over to their shifted counterparts.

References

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

[2] 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.

[3] L. Ekenberg, M. Andersson, M. Danielson, and A. Larsson. Distributions over Expected Utilities in Decision Analysis. *Proceedings of ISIPTA '07*, 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] E.T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, 1957.

[7] E.T. Jaynes. Information Theory and Statistical Mechanics II. *Physical Review*, 108(2):171–190, 1957.

[8] S. Kullback and R.A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

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

[10] A. W. Marshall and I. Olkin. Families of Multivariate Distributions. *Journal of the American Statistical Assoiciation*, 83(403):834–841, September 1988.

[11] R. F. Nau. Uncertainty Aversion with Second-Order Utilities and Probabilities. *Management Science*, 52(1):136–145, 2006.

[12] R. B. Nelsen. *An introduction to copulas, Volume 139 of Lecture Notes in Statistics*. Springer Verlag, 1999.

[13] D. Sundgren, M. Danielson, and L. Ekenberg. Some Second Order Effects on Interval Based Probabilities. In *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference*, pages 848–853. AAAI Press, 2006.

[14] L. 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.

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

[16] P. Walley. Towards a Unified Theory of Imprecise Probability. *International Journal of Approximate Reasoning*, 24(2-3):125–148, 2000.

[17] S. Watanabe. Infomation Theoretical Analysis of Multivariate Correlation. *IBM Journal of Research and Development*, 4(1):66–82, 1960.