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

David Sundgren, Love Ekenberg and Mats Danielson Royal Institute of Technology, Sweden

KTH Information and Communication Technology

Abstract

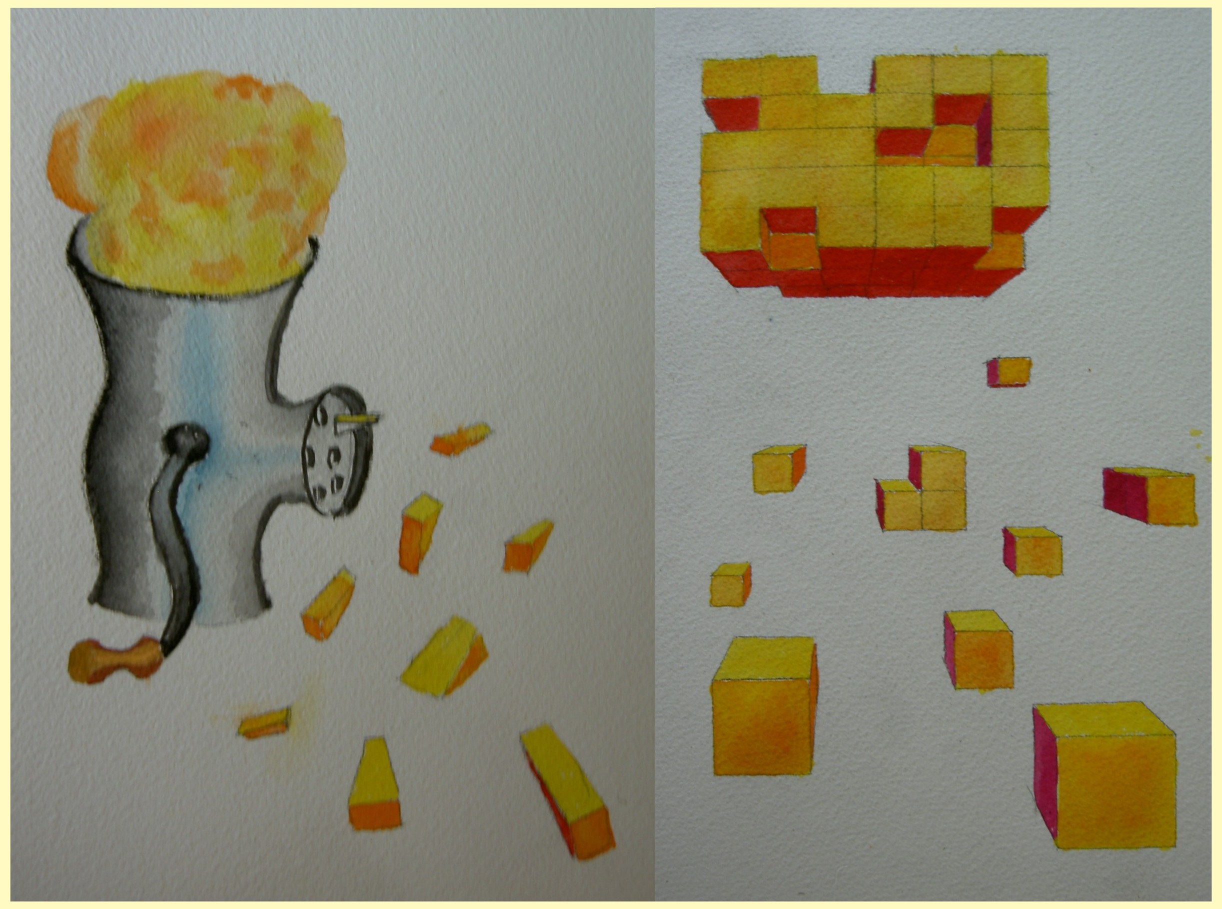
In classic decision theory it is assumed that a decision-maker can assign precise numerical values corresponding to the true value of each consequence, as well as precise numerical probabilities for their occurrences. In attempting to address real-life problems, where uncertainty in the input data prevails, 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.

Second-Order Distributions

Uncertain probabilities are thought to be represented by 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 $1 - \sum_{i=1}^{n-1} x_i$.
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.

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

Where is the Uncertainty?



Say that we do not know anything except lower and upper bounds for the probabilities and want to employ the principle of maximum entropy. Do we let all the entropy reside in the multivariate (global) second-order distribution or do we distribute our uncertainty on both the global and local levels?

Factoring into Marginals

We want to find multivariate distributions 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 $K = \int_{\mathcal{P}} \prod_{i=1}^{n} f_i(x_i)\, d\mathbf{x}$.
The idea is that the entropy is found both at the global and local levels, or that the *Kullback-Leibler divergence* $D_{\mathrm{KL}}(f \,\|\, \prod_{i=1}^{n} f_i)$, also known as the *total correlation* of $X_1, \ldots, X_n$ to be minimal. We could even say that we wish the first-order probabilities to be as close to independent as possible given that the probabilities sum to one.

Dirichlet Rediscovered

The distributions we seek turns out be Dirichlet distributions with parameters $\alpha_i = \frac{1}{n-1}$ and the marginal distributions Beta distributions with $\alpha = \frac{1}{n-1}$ and $\beta = 1$ when all lower bounds $a_i = 0$.
When any $a_i > 0$ the distributions are shifted and re-scaled according to the new support.
The upper bound of x_j is determined by the lower bounds of the other variables, $1 - \sum_{i \neq j} a_i$.

The Theorem

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 $\sum_{j=1}^{n} a_j < 1$.

Proof: 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 $i, i = 1, \ldots, n$,

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

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

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

Example

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 $[1/8, 7/15]$.