Overdrawing Urns using Categories of Signed Probabilities

Bart Jacobs and Dario Stein
iHub, Radboud University Nijmegen

July 6, 2023

bart.jacobs@ru.nl dario.stein@ru.nl

A basic experiment in probability theory is drawing without replacement from an urn filled with multiple balls of different colours. Clearly, it is physically impossible to overdraw, that is, to draw more balls from the urn than it contains. This paper demonstrates that overdrawing does make sense mathematically, once we allow signed distributions with negative probabilities. A new (conservative) extension of the familiar hypergeometric ('draw-and-delete') distribution is introduced that allows draws of arbitrary sizes, including overdraws. The underlying theory makes use of the dual basis functions of the Bernstein polynomials, which play a prominent role in computer graphics. Negative probabilities are treated systematically in the framework of categorical probability and the central role of datastructures such as multisets and monads is emphasised.

1 Introduction

For drawing (multiple) coloured balls from a statistical urn, we distinguish three well-known modes:

1. *hypergeometric* or *draw-and-delete*, which is drawing a ball from the urn without replacement, so that the urn shrinks;
2. *multinomial* or *draw-and-replace*: drawing with replacement, so that the urn remains the same;
3. *Pólya* or *draw-and-duplicate*, which is drawing a ball from the urn and replacing it together with an additional ball of the same colour, so that the urn grows.

Multinomial and Pólya draws may be of arbitrary size, but hypergeometric draws are limited in size by the number of balls in the urn. In this paper we lift this limitation and allow hypergeometric draws of arbitrary size, including 'overdraws', containing more balls than in the urn. Physically this is strange, but, as will show, mathematically it makes sense once we allow negative probabilities.

Negative probabilities have emerged in quantum physics (*e.g.* in double slit experiments) and have been discussed in the work of famous physicists like Wigner, Dirac, and Feynman (see *e.g.* [7] and the references mentioned there). There are also 'classical' (non-quantum) examples, such as the one of Piponi (discussed in [1]) or of Székely [22] with two half coins, involving infinitely many both positive and negative probabilities, whose (convolution) sum is an ordinary (fair) coin. Also, negative probabilities have come up in finance, see *e.g.* [20]. Despite the lack of the clear operational meaning that their nonnegative counterparts have, negative probabilities appear as convenient tools in a variety of contexts in mathematics and physics, see [2, 23, 1, 7].

We briefly explain the nature of our extension, already using some notation that will be explained below. The Pólya distribution can be expressed as a mixture of multinomial draws; that is, we can break up such a draw into two stages: first sample a random distribution ω from the Dirichlet distribution, and then make independent (multinomial) draws from ω. The self-reinforcing behaviour of Pólya's urn is entirely captured by the latent Dirichlet distribution. In the Kleisli category of the Giry-monad,

Submitted to:
ACT 2023

with (Kleisli) morphisms called channels, this corresponds to a factorisation of the Pólya channel $pol[K]$ through the multinomial channel $mn[K]$, with draws of size K.

$$pol[K] \quad = \quad \mathcal{M}[N](X) \xrightarrow{Dir} \mathcal{D}(X) \xrightarrow{mn[K]} \mathcal{M}[K](X) \tag{1}$$

We write $\mathcal{M}[N](X)$ for the space of multisets (urns) on a set X of size N, and $\mathcal{D}(X)$ is the set of finite distributions. Since such factorisations arise from De Finetti's famous theorem, we call (1) a De Finetti factorisation for the Pólya's distribution (*e.g.* [15]).

The hypergeometric distribution does not admit such a De Finetti factorisation since an urn containing N balls is exhausted after N draws. However, there is a way out, if we extend our notion of probability to allow negative (signed) probabilities. Such models satisfy the usual axioms of categorical probability, and we can find a De Finetti factorization of the hypergeometric channel $hg[K]$, for draws of size K, of the form:

$$hg[K] \quad = \quad \mathcal{M}[N](X) \xrightarrow{DDir} \mathcal{D}(X) \xrightarrow{mn[K]} \mathcal{M}[K](X) \tag{2}$$

It uses a signed 'Dual Dirichlet' distribution $DDir$ which we develop in analogy to the Dirichlet distribution occurring in Pólya's urn. Existence of the factorisation can be deduced from earlier work [16, 5, 18] connecting finite versions of De Finetti's theorem to signed probability. In this case, the factorisation (2) is not unique. We claim that a canonical choice is given by the Dual Bernstein polynomials, which have been studied widely in computer graphics [19, 6, 24, 17], but their appearance in a probabilistic context is novel. Evaluating (2) for overdraws $K \geq N$ defines a signed extension of the hypergeometric distributions which includes overdraws, while agreeing with the usual distribution for ordinary draws $K \leq N$.

Our contributions are:

1. a principled approach to signed probability (discrete and continuous) using multisets and monads;
2. conceptualizing dual Bernstein polynomials as signed probability densities;
3. defining signed hypergeometric distributions that conservatively extend hypergeometric draws while preserving good properties;
4. explicating the dual Dirichlet distribution and its conjugate prior relationships via string diagrams.

2 Multisets

A multiset, also known as bag, is like a subset except that elements may occur multiple times. We shall use ket notation $n_1|x_1\rangle + \cdots + n_k|x_k\rangle$ to describe a multiset with k elements, where element x_i, say from a set X, occurs $n_i \in \mathbb{N}$ many times. Equivalently, such a multiset may be described as a function $\varphi \colon X \to \mathbb{N}$ with finite support $supp(\varphi) \coloneqq \{x \in X \mid \varphi(x) \neq 0\}$. The number $\varphi(x) \in \mathbb{N}$ is the multiplicity of $x \in X$; it says how many times x occurs in the multiset φ.

We shall write $\mathcal{M}(X)$ for the set of multisets with elements from a set X, and $\mathcal{M}_{\mathrm{fs}}(X) \subseteq \mathcal{M}(X)$ for the subset of multisets φ with full support, that is, with $supp(\varphi) = X$. The latter only makes sense when X is a finite set. As canonical finite sets we write $\boldsymbol{n} \coloneqq \{0, 1, \ldots, n-1\}$, for $n \in \mathbb{N}$.

The size $\|\varphi\|$ of a multiset φ is the total number of elements, including multiplicities. Thus, $\|\varphi\| \coloneqq \sum_x \varphi(x)$, or, in ket notation, $\|\sum_i n_i|x_i\rangle\| = \sum_i n_i$. We write $\mathcal{M}[K](X) \coloneqq \big\{\varphi \in \mathcal{M}(X) \;\big|\; \|\varphi\| = K\big\}$ for the set of multisets of size $K \in \mathbb{N}$. When the set X has $n \geq 1$ elements, the number of multisets of size K in $\mathcal{M}[K](X)$ is $\left(\!\binom{n}{K}\!\right) = \binom{n+K-1}{K} = \frac{(n+K-1)!}{K!\cdot(n-1)!}$. For instance, for a set $X = \{a,b,c\}$ with three elements there are $\left(\!\binom{3}{3}\!\right) = \frac{5!}{3!\cdot 2!} = 10$ multisets of size $K = 3$, namely: $3|a\rangle$, $3|b\rangle$, $3|c\rangle$, $2|a\rangle + 1|b\rangle$, $2|a\rangle + 1|c\rangle$,

$1|a\rangle + 2|b\rangle$, $2|b\rangle + 1|c\rangle$, $1|a\rangle + 2|c\rangle$, $1|b\rangle + 2|c\rangle$, $1|a\rangle + 1|b\rangle + 1|c\rangle$. Only the last one has full support. The factorial $n!$ and binomial coefficients $\binom{n}{m}$ and $\left(\!\binom{n}{m}\!\right)$ are extended from numbers to multisets (as in [13]).

Definition 1 *Let $\varphi, \psi \in \mathcal{M}(X)$ be two multisets. We define*

1. $\varphi\text{‼} := \prod_x \varphi(x)!$;
2. $\boldsymbol{(}\varphi\boldsymbol{)} := \frac{\|\varphi\|!}{\varphi\text{‼}}$;
3. $\varphi \leq \psi$ *iff* $\varphi(x) \leq \psi(x)$ *for each* $x \in X$, *and* $\varphi \leq_K \psi$ *iff* $\varphi \leq \psi$ *and* $\|\varphi\| = K$;
4. $(\psi - \varphi)(x) = \psi(x) - \varphi(x)$, *when* $\varphi \leq \psi$;
5. $\binom{\psi}{\varphi} := \frac{\psi\text{‼}}{\varphi\text{‼} \cdot (\psi - \varphi)\text{‼}} = \prod_x \binom{\psi(x)}{\varphi(x)}$, *when* $\varphi \leq \psi$;
6. $\left(\!\binom{\psi}{\varphi}\!\right) := \frac{(\psi + \varphi - \mathbf{1})\text{‼}}{\varphi\text{‼} \cdot (\psi - \mathbf{1})\text{‼}} = \prod_x \left(\!\binom{\psi(x)}{\varphi(x)}\!\right)$ *when* ψ *has full support, where* $\mathbf{1} = \sum_x 1|x\rangle$ *is the multiset of singletons.*

3 Discrete distributions

A discrete probability distribution $\sum_i r_i|x_i\rangle$ looks like a multiset, except that the multiplicities r_i are now in the unit interval $[0,1] \subseteq \mathbb{R}$ and add up to one: $\sum_i r_i = 1$. We write $\mathcal{D}(X)$ for the set of such distributions with $x_i \in X$. Alternatively, like for multisets, elements $\omega \in \mathcal{D}(X)$ may be described as functions $\omega \colon X \to [0,1]$ with finite support and with $\sum_x \omega(x) = 1$. When the set X is finite, we write $\mathcal{D}_{\text{fs}}(X) \subseteq \mathcal{D}(X)$ for the discrete distributions with full support: $\mathit{supp}(\omega) = X$. An example is the uniform distribution $\sum_{x \in X} \frac{1}{n}|x\rangle$, where $n \geq 1$ is the number of elements of a non-empty set X. Concretely, a fair coin is described by the distribution $\frac{1}{2}|H\rangle + \frac{1}{2}|T\rangle$ for $X = \{H, T\}$.

Each non-empty multiset $\varphi \in \mathcal{M}(X)$ can be turned into a distribution via normalisation. We call this frequentist learning, since it involves learning a distribution by counting, and write it as:

$$\mathit{flrn}(\varphi) := \frac{\varphi}{\|\varphi\|} = \sum_{x \in X} \frac{\varphi(x)}{\|\varphi\|}\,\big|x\big\rangle \in \mathcal{D}(X). \tag{3}$$

This frequentist learning is natural in X, but it is not a map of monads, from (non-empty) multisets to distributions.

The set $\mathcal{D}(\boldsymbol{n})$ of distributions on $\boldsymbol{n} = \{0, \ldots, n-1\}$ can be identified with the simplex $\Delta^n \subseteq \mathbb{R}^n$, where:

$$\Delta^n := \big\{(r_0, \ldots, r_{n-1}) \in \mathbb{R}_{\geq 0} \;\big|\; \textstyle\sum_i r_i = 1\big\}. \tag{4}$$

This is commonly called the $n-1$ simplex.

There are three famous 'draw' distributions, called multinomial, hypergeometric and Pólya. We briefly describe them in the style of [14] and refer there for more information. These distributions are all parameterised by a draw size K and form distributions on the set $\mathcal{M}[K](X)$ of multisets (as draws). One may think of X as a set of colors.

Definition 2 *We fix a set X and a number $K \in \mathbb{N}$.*

1. *For a distribution $\omega \in \mathcal{D}(X)$, used as abstract urn, the multinomial distribution $mn[K](\omega) \in \mathcal{D}\big(\mathcal{M}[K](X)\big)$ is defined as:*

$$mn[K](\omega) := \sum_{\varphi \in \mathcal{M}[K](X)} \boldsymbol{(}\varphi\boldsymbol{)} \cdot \prod_{x \in X} \omega(x)^{\varphi(x)}\,\big|\varphi\big\rangle.$$

2. *For an 'urn' multiset $\upsilon \in \mathcal{M}(X)$, with size $L := \|\upsilon\| \geq K$ there is a hypergeometric distribution $hg[K](\upsilon) \in \mathcal{D}\big(\mathcal{M}[K](X)\big)$ with:*

$$hg[K](\upsilon) := \sum_{\varphi \leq_K \upsilon} \frac{\binom{\upsilon}{\varphi}}{\binom{L}{K}} \big| \varphi \big\rangle.$$

The size restriction $K \leq L$ excludes overdraws.

3. *Similarly, for a multiset $\upsilon \in \mathcal{M}[L](X)$ with full support, there is the Pólya distribution $pol[K](\upsilon) \in \mathcal{D}\big(\mathcal{M}[K](X)\big)$ with:*

$$pol[K](\upsilon) := \sum_{\varphi \in \mathcal{M}[K](X)} \frac{\left(\!\binom{\upsilon}{\varphi}\!\right)}{\left(\!\binom{L}{K}\!\right)} \big| \varphi \big\rangle.$$

Intuitively, in the multinomial case drawn balls are returned to the urn, so that the urn does not change and can be described abstractly as a discrete distribution. In the hypergeometric case a drawn ball is removed from the urn, and in the Pólya case the drawn ball is returned together with a new ball of the same colour. Thus in the hypergeometric case the urns shrinks, whereas in the Pólya case the urn grows.

Given two distributions $\omega \in \mathcal{D}(X)$ and $\rho \in \mathcal{D}(Y)$ we can form a parallel (tensor) product $\omega \otimes \rho \in \mathcal{D}(X \times Y)$, via $(\omega \otimes \rho)(x,y) = \omega(x) \cdot \rho(y)$.

The mapping $X \mapsto \mathcal{D}(X)$ is a monad on the category of sets. We will not spell out what this means, but we will use the resulting Kleisli category $\mathcal{K}\ell(\mathcal{D})$, whose maps $c \colon X \to \mathcal{D}(Y)$ will be called channels and written as $c \colon X \rightsquigarrow Y$. For instance, the distributions from Definition 2 can be described as channels $mn[K] \colon \mathcal{D}(X) \rightsquigarrow \mathcal{M}[K](X)$, $hg[K] \colon \mathcal{M}[L](X) \rightsquigarrow \mathcal{M}[K](X)$ and $pol[K] \colon \mathcal{M}_{\mathrm{fs}}(X) \rightsquigarrow \mathcal{M}[K](X)$.

For a channel $c \colon X \rightsquigarrow Y$ and a distribution $\omega \in \mathcal{D}(X)$ on the domain X we can form a distribution $c \gg\!= \omega$ on the codomain Y via pushforward (also called state transformation):

$$\big(c \gg\!= \omega\big)(y) := \sum_{x \in X} \omega(x) \cdot c(x)(y).$$

Given another channel $d \colon Y \rightsquigarrow Z$ one can form a composite channel $d \circ\!\!\cdot\, c \colon X \rightsquigarrow Z$ via $(d \circ\!\!\cdot\, c)(x) := d \gg\!= c(x)$. Notice that we use a special circle $\circ\!\!\cdot$, with a dot, for composition of channels.

A basic channel is $DD \colon \mathcal{M}[K{+}1](X) \rightsquigarrow \mathcal{M}[K](X)$, where DD stands for draw-delete. It probabilistically draws and removes one ball from an urn $\upsilon \in \mathcal{M}[K{+}1](X)$ with $K{+}1$ balls, via:

$$DD(\upsilon) := \sum_{x \in supp(\upsilon)} \frac{\upsilon(x)}{\|\upsilon\|} \big| \upsilon - 1|x\rangle \big\rangle. \tag{5}$$

We recall, without proof, the following basic properties of draw distributions, mostly from [13], expressed in terms of channels.

Proposition 3 *1.* $flrn \circ\!\!\cdot\, hg[K] = flrn$;

2. $flrn \circ\!\!\cdot\, mn[K] = sam$, *where* $sam \colon \mathcal{D}(X) \rightsquigarrow X$ *is the identity map, considered as channel;*
3. $hg[K] \circ\!\!\cdot\, mn[K{+}L] = mn[K]$;
4. $hg[K] \circ\!\!\cdot\, hg[K{+}L] = hg[K]$;
5. $hg[K] \circ\!\!\cdot\, DD = hg[K]$;

6. $DD \circ hg[K{+}1] = hg[K]$;
7. $DD \circ mn[K{+}1] = mn[K]$;
8. $DD \circ pol[K{+}1] = pol[K]$. ◀

The last two items express that multinomial and Pólya form cones for the infinite chain of draw-delete channels that appears in a categorical perspective on De Finetti's theorem, see [15] and [16]. This fails in the hypergeometric case since the draw size K must remain smaller than the size of the urn.

4 Continuous distributions

In the previous section, we have seen finite discrete probability distributions over an arbitrary set. There are also continuous distributions, defined on measurable spaces. Here we need such distributions only on one particular kind of spaces, namely on simplices Δ^n, see (4). The only distributions that we need are given by (polynomial) functions $f\colon \Delta^n \to \mathbb{R}$ with $\int_{\Delta^n} f = 1$. Such an f is called a (probability) density function. It gives rise to probability measure Φ that sends a measurable subset $M \subseteq \Delta^n$ to the probability $\int_M f \in [0,1]$. Such measures are elements of the set $\mathcal{G}(\Delta^n)$, where $\mathcal{G}$ is the Giry monad, see *e.g.* [21, 11] for further information. At first we require that such density functions are nonnegative, so $f \geq 0$, but later we drop this requirement, for so-called signed distributions (see the next section).

Definition 4 *Let $\upsilon \in \mathcal{M}_{\mathrm{fs}}(\boldsymbol{n})$ be an urn with full support (for $n \geq 1$).*

1. *It gives rise to the Dirichlet density $dir(\upsilon)\colon \Delta^n \to \mathbb{R}_{\geq 0}$ given on $\boldsymbol{r} \in \Delta^n$ by:*

$$dir(\upsilon)(\boldsymbol{r}) := \frac{(\|\upsilon\|-1)!}{(\upsilon - \mathbf{1}_n)\underset{\circ}{!}} \cdot \prod_{i\in\boldsymbol{n}} r_i^{\upsilon(i)-1} \qquad \textit{where} \qquad \mathbf{1}_n := \sum_{0\leq i<n} 1|\,i\,\rangle \in \mathcal{M}_{\mathrm{fs}}(\boldsymbol{n}).$$

   We shall drop the index n from $\mathbf{1}_n$ when it is clear from the context.

2. *The associated probability measure $Dir(\upsilon)$ is defined on measurable subsets $M \subseteq \Delta^n$ as:*

$$Dir(\upsilon)(M) := \int_{\boldsymbol{r}\in M} dir(\upsilon)(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}.$$

The function $dir(\upsilon)$ in (1) is a proper probability density because of the following standard equation that explains the form of the Dirichlet normalisation constant, for $\upsilon \in \mathcal{M}_{\mathrm{fs}}(\boldsymbol{n})$.

$$\int_{\boldsymbol{r}\in\Delta^n} \prod_{i\in\boldsymbol{n}} r_i^{\upsilon(i)-1}\,\mathrm{d}\boldsymbol{r} = \frac{(\upsilon-\mathbf{1})\underset{\circ}{!}}{(\|\upsilon\|-1)!}. \tag{6}$$

We use Dirichlet for urns υ with positive natural numbers as multiplicities. This can be generalised to urns with positive real numbers as multiplicities — using the Gamma function instead of factorials — but that is not needed in the current setting.

In the sequel we shall use the bind notation $c \gg\!= \Phi$ also for continuous measures, but in a very restricted form, namely for measures Φ on Δ^n given by a probability density function f and for channels $c\colon \mathcal{D}(\boldsymbol{n}) \rightsquigarrow \mathcal{M}[K](\boldsymbol{n})$. Categorically, this bind is the Kleisli extension for the Giry monad $\mathcal{G}$, see *e.g.* [21, 8] for details. We will not elaborate this background and will simply use the relevant equation, which is of the following form, for $\varphi \in \mathcal{M}[K](\boldsymbol{n})$,

$$c \gg\!= \Phi := \sum_{\varphi\in\mathcal{M}[K](\boldsymbol{n})} \left(\int_{\boldsymbol{r}\in\Delta^n} c(\boldsymbol{r})(\varphi)\cdot f(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\right)\big|\,\varphi\,\big\rangle, \qquad \text{where we identify } \mathcal{D}(\boldsymbol{n}) \text{ and } \Delta^n. \tag{7}$$

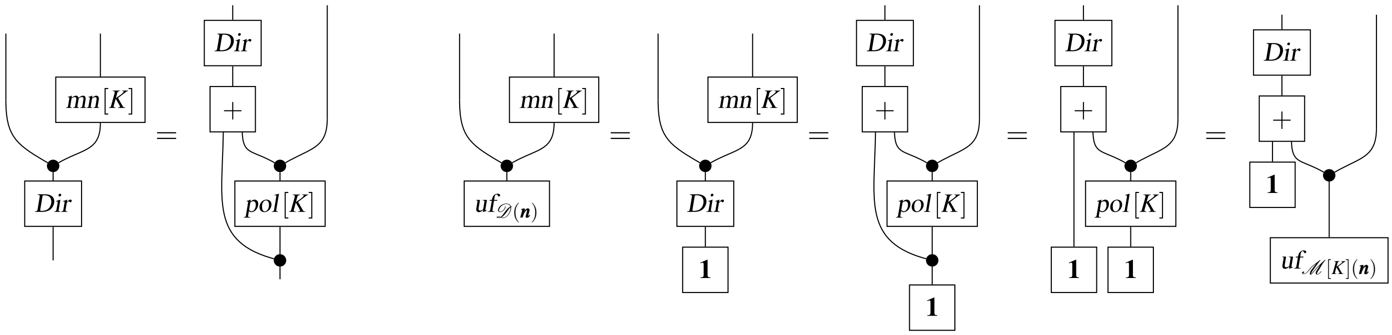


Figure 1: On the left the equation expressing that multinomial is a sufficient statistic for Dirichlet; on the right the string diagrammatic proof that the dagger of Dirichlet is multinomial, with the uniform distribution as prior, see Theorem 5 for details. The boxed **1** is the point/singleton distribution $1|\mathbf{1}\rangle$. Such point distributions commute with copiers.

The next result summarises the close relationship between multinomial, Pólya, and Dirichlet distributions. The first two points are well-known, but the third one probably a bit less — although it follows easily from the conjugate prior situation (see also [12]). We use string diagrammatic notation, with flows from bottom to top, since it best displays what is going on, see [8] for details. Proofs are in Appendix B.

Theorem 5 *Let $\upsilon \in \mathcal{M}_{\mathrm{fs}}(\boldsymbol{n})$ be a multiset / urn with full support, of size $L := \|\upsilon\|$, where $n \geq 1$, and let K be an arbitrary number.*

1. *Multinomial over Dirichlet is Pólya: $mn[K] \gg\!= Dir(\upsilon) = pol[K](\upsilon)$.*
2. *Multinomial is conjugate prior of Dirichlet: updating $Dir(\upsilon)$ with the predicate / likelihood $mn[K](-)(\varphi)$ is $Dir\big(\upsilon + \varphi\big)$. This is expressed diagrammatically on the left in Figure 1.*
3. *Multinomial is the dagger of Dirichlet w.r.t. the uniform distribution $uf_{\mathcal{D}(\boldsymbol{n})}$ on $\mathcal{D}(\boldsymbol{n})$. In the language / notation of [4, 3] this is expressed as: $mn[K] = Dir\big(\mathbf{1} + -\big)^{\dagger}_{uf_{\mathcal{D}(\boldsymbol{n})}}$.*
4. *When we slightly massage the sample channel from Proposition 3 (2) to $sam\colon \Delta^n \rightsquigarrow \boldsymbol{n}$ given by $sam(\boldsymbol{r}) = \sum_{i\in\boldsymbol{n}} r_i|i\rangle$, then: $sam \gg\!= Dir(\upsilon) = flrn(\upsilon)$.* ◀

The first item of Theorem 5 tells that Pólya is multinomial over Dirichlet. This is an important starting point for this paper, since we asked ourselves the question whether there is also a distribution, like Dirichlet, such that multinomial over it is hypergeometric. We shall see below that the so-called 'signed' Dirichlet distributions achieve this. But first we need to set the scene for these signed distributions.

5 Signed distributions

We now introduce signed distribution, both in the discrete case and in the continuous case. As before, we only need continuous distributions on simplices.

Definition 6 1. *A signed discrete probability distribution on a set X is a function $\sigma\colon X \to \mathbb{R}$ with finite support $supp(\sigma) := \{x \in X \mid \sigma(x) \neq 0\}$ and with $\sum_{x\in X} \sigma(x) = 1$. We may equivalently write such a signed discrete distribution in ket notation as a finite formal sum $\sum_i r_i|x_i\rangle$ where $r_i \in \mathbb{R}$ satisfy $\sum_i r_i = 1$. We shall write $\mathcal{S}(X)$ for the set of signed discrete probability distributions on X.*

2. *A signed continuous probability distribution, on a simplex Δ^n, is given by a signed density function $f\colon \Delta^n \to \mathbb{R}$ with $\int_{\boldsymbol{r}\in\Delta^n} f(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = 1$.*

An example of a signed discrete distribution is $\frac{1}{2}|a\rangle - \frac{1}{4}|b\rangle + \frac{3}{4}|c\rangle$. We do not offer an operational explanation for what such negative probabilities mean but treat signed distributions as mathematical objects of their own. It is not hard to see that signed discrete distributions $\mathscr{S}$ form a monad on the category of sets and functions. It is affine, in the sense that $\mathscr{S}(\mathbf{1}) \cong \mathbf{1}$, but it differs from $\mathscr{D}$ for instance because it is not *strongly* affine, as defined in [10].

6 Dual bases

The probability mass function of the multinomial distribution is of a particularly tractable form, namely a polynomial function $\Delta^n \to \mathbb{R}$, on the simplex Δ^n.

Definition 7 *For $\varphi \in \mathscr{M}[K](\boldsymbol{n})$, we define the* multinomial $\mathfrak{m}_\varphi$ *as*

$$\mathfrak{m}_\varphi(\boldsymbol{x}) \coloneqq (\varphi) \cdot \boldsymbol{x}^\varphi = (\varphi) \cdot \prod_{i \in \boldsymbol{n}} x_i^{\varphi(i)} \qquad \textit{with 'monomial'} \qquad \boldsymbol{x}^\varphi \coloneqq \prod_{i \in \boldsymbol{n}} x_i^{\varphi(i)}.$$

For every probability vector $\boldsymbol{r} \in \Delta^n$, one has $\mathfrak{m}_\varphi(\boldsymbol{r}) = mn[K](\boldsymbol{r})(\varphi)$, via the identification $\Delta^n \cong \mathscr{D}(\boldsymbol{n})$.

Definition 8 *For numbers n, K we write $P_K(\Delta^n)$ for the real vector space of polynomial functions $\Delta^n \to \mathbb{R}$ of degree K. The multinomials $\mathfrak{m}_\varphi$ for $\varphi \in \mathscr{M}[K](\boldsymbol{n})$ form a basis of this space, and so we have as dimension* $\dim\big(P_K(\Delta^n)\big) = \left(\!\binom{n}{K}\!\right)$. *This vector space $P_K(\Delta^n)$ is a Hilbert space via an inner product defined on $f, g \colon \Delta^n \to \mathbb{R}$ as:*

$$\big\langle f, g \big\rangle \coloneqq \int_{\boldsymbol{r} \in \Delta^n} f(\boldsymbol{r}) \cdot g(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r}. \tag{8}$$

The dual of a basis (b_i) of a space V is generally understood as a basis of the dual space V^*. In a Hilbert space such a dual basis can be described as the elements (d_i) of the space itself which are uniquely determined by the relationship $\langle b_i, d_j \rangle = \delta_{ij}$, so that $\langle b_i, d_i \rangle = 1$ and $\langle b_i, d_j \rangle = 0$ for $i \neq j$.

Definition 9 *The* dual multinomials $(\mathfrak{d}_\varphi)$ *are defined as the dual basis of $P_K(\Delta^n)$ to the multinomials $(\mathfrak{m}_\varphi)$, and are as such uniquely characterised by the property $\varphi, \psi \in \mathscr{M}[K](\boldsymbol{n})$,*

$$\big\langle \mathfrak{m}_\varphi, \mathfrak{d}_\psi \big\rangle = \delta_{\varphi,\psi} = \begin{cases} 1 & \textit{if } \varphi = \psi \\ 0 & \textit{if } \varphi \neq \psi, \end{cases} \tag{9}$$

What do we know about this dual basis? Of course we can express the dual basis vectors $\mathfrak{d}_\psi$ in terms of the original basis, say via scalars $c_{\chi,\psi}$ satisfying, for each $\psi \in \mathscr{M}[K](\boldsymbol{n})$,

$$\mathfrak{d}_\psi = \sum_{\chi \in \mathscr{M}[K](\boldsymbol{n})} c_{\chi,\psi} \cdot \boldsymbol{x}^\chi = \sum_{\chi \in \mathscr{M}[K](\boldsymbol{n})} \frac{c_{\chi,\psi}}{(\chi)} \cdot \mathfrak{m}_\chi. \tag{10}$$

By exploiting the equations (9) and using the linearity of the inner product in each of its arguments (*i.e.* bilinearity), we obtain the equation

$$\delta_{\varphi,\psi} = \big\langle \mathfrak{m}_\varphi, \mathfrak{d}_\psi \big\rangle = \sum_{\chi \in \mathscr{M}[K](\boldsymbol{n})} c_{\chi,\psi} \cdot \frac{\big\langle \mathfrak{m}_\varphi, \mathfrak{m}_\chi \big\rangle}{(\chi)}. \tag{11}$$

We note that:

$$\begin{aligned} \big\langle \mathfrak{m}_\varphi, \mathfrak{m}_\chi \big\rangle &\overset{(8)}{=} (\varphi) \cdot (\chi) \cdot \int_{\boldsymbol{r} \in \Delta^n} \left(\prod_{i \in \boldsymbol{n}} r_i^{\varphi(i)} \right) \cdot \left(\prod_{i \in \boldsymbol{n}} r_i^{\chi(i)} \right) \mathrm{d}\boldsymbol{r} = (\varphi) \cdot (\chi) \cdot \int_{\boldsymbol{r} \in \Delta^n} \prod_{i \in \boldsymbol{n}} r_i^{(\varphi+\chi)(i)} \, \mathrm{d}\boldsymbol{r} \\ &\overset{(6)}{=} (\varphi) \cdot (\chi) \cdot \frac{(\varphi+\chi)!}{(2K+n-1)!} \end{aligned}$$

There are three square matrices at hand, of size $\left(\!\binom{n}{K}\!\right) \times \left(\!\binom{n}{K}\!\right)$, with multisets as indices, namely:

$$C = \Big(c_{\varphi,\psi}\Big)_{\varphi,\psi\in\mathcal{M}[K](\boldsymbol{n})} \qquad FS = \Big((\varphi+\psi)!\Big)_{\varphi,\psi\in\mathcal{M}[K](\boldsymbol{n})} \qquad D = \Big((\varphi)\cdot\delta_{\varphi,\psi}\Big)_{\varphi,\psi\in\mathcal{M}[K](\boldsymbol{n})}.$$

This C is the matrix of scalars that we are looking for, FS contains the factorials-of-sums of multisets, and D is a diagonal matrix with multiset coefficients. Equation (11) can now be written as:

$$(2K{+}n{-}1)!\cdot\delta_{\varphi,\psi} = \sum_{\chi\in\mathcal{M}[K](\boldsymbol{n})} (\varphi)\cdot FS_{\varphi,\chi}\cdot C_{\chi,\psi} = \Big(D\cdot FS\cdot C\Big)_{\varphi,\psi}.$$

We then get $(2K+n-1)!\cdot FS^{-1}\cdot D^{-1} = C$, so that the coefficients that we seek are obtained as:

$$c_{\varphi,\psi} = \frac{(2K{+}n{-}1)!}{(\psi)}\cdot\Big(FS^{-1}\Big)_{\varphi,\psi}. \tag{12}$$

These matrix inverses FS^{-1} exist since FS is a symmetric positive definite matrix. We give an extended example calculation in the appendix (Example 19).

The following is a crucial property of dual bases, as introduced in Definition 9.

Proposition 10 *Each dual basis function $\mathfrak{d}_\psi \in P_K(\Delta^n)$, associated with a multiset $\psi\in\mathcal{M}[K](\boldsymbol{n})$, is a continuous signed probability density on Δ^n, that is:*

$$\int_{\boldsymbol{r}\in\Delta^n} \mathfrak{d}_\psi(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = 1.$$

Proof We use that multinomial distributions $mn[K]$ form a probability distribution; this means that the multinomials $\mathfrak{m}_\varphi$ form a partition of unity, *i.e.* for all $\boldsymbol{r}\in\Delta^n$:

$$\sum_{\varphi\in\mathcal{M}[K](\boldsymbol{n})} \mathfrak{m}_\varphi(\boldsymbol{r}) = 1.$$

Hence we obtain

$$\begin{aligned}
\int_{\boldsymbol{r}\in\Delta^n} \mathfrak{d}_\psi(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = \int_{\boldsymbol{r}\in\Delta^n} 1\cdot\mathfrak{d}_\psi(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} &= \int_{\boldsymbol{r}\in\Delta^n} \Bigg(\sum_{\varphi\in\mathcal{M}[K](\boldsymbol{n})} \mathfrak{m}_\varphi(\boldsymbol{r})\Bigg)\cdot\mathfrak{d}_\psi(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \\
&= \sum_{\varphi\in\mathcal{M}[K](\boldsymbol{n})} \int_{\boldsymbol{r}\in\Delta^n} \mathfrak{m}_\varphi(\boldsymbol{r})\cdot\mathfrak{d}_\psi(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \\
&\overset{(8)}{=} \sum_{\varphi\in\mathcal{M}[K](\boldsymbol{n})} \big\langle \mathfrak{m}_\varphi, \mathfrak{d}_\psi \big\rangle \overset{(9)}{=} \sum_{\varphi\in\mathcal{M}[K](\boldsymbol{n})} \delta_{\varphi,\psi} = 1. \qquad \blacktriangleleft
\end{aligned}$$

In the beginning of this proof, we use that the pointwise sum of the multinomial basis functions $\mathfrak{m}_\varphi$, for $\varphi\in\mathcal{M}[K](\boldsymbol{n})$, is equal to the constant-one function $\mathbf{1}\colon\Delta^n\to\mathbb{R}$. The sum of the dual basis functions $\mathfrak{d}_\varphi$ is also constant.

Proposition 11 *Fix numbers n,K and consider the dual basis function $\mathfrak{d}_\varphi\in P_K(\Delta^n)$, for $\varphi\in\mathcal{M}[K](\boldsymbol{n})$. Then their pointwise sum is a constant function:*

$$\sum_{\varphi\in\mathcal{M}[K](\boldsymbol{n})} \mathfrak{d}_\varphi = \frac{(K+n-1)!}{K!}.$$

Proof Since the vectors $\mathfrak{d}_\varphi$ form a basis we can express the constant-one function $\mathbf{1}\colon \Delta^n \to \mathbb{R}$ with respect to this basis, say as: $\mathbf{1} = \sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})} a_\varphi \cdot \mathfrak{d}_\varphi$, for certain coefficients a_φ. For a fixed multiset $\psi \in \mathscr{M}[K](\boldsymbol{n})$ we compute the constant a_ψ as follows.

$$\begin{aligned}
a_\psi \;&=\; \sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})} a_\varphi \cdot \delta_{\psi,\varphi} \;=\; \sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})} a_\varphi \cdot \langle \mathfrak{m}_\psi, \mathfrak{d}_\varphi \rangle \;=\; \Big\langle \mathfrak{m}_\psi, \sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})} a_\varphi \cdot \mathfrak{d}_\varphi \Big\rangle \\
&=\; \langle \mathfrak{m}_\psi, \mathbf{1} \rangle \;=\; \int_{\boldsymbol{r}\in\Delta^n} \mathfrak{m}_\psi(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \;=\; \boldsymbol{(}\psi\boldsymbol{)} \cdot \int_{\boldsymbol{r}\in\Delta^n} \boldsymbol{r}^\psi(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \overset{(6)}{=} \frac{K!}{\psi!} \cdot \frac{\psi!}{(K+n-1)!} \;=\; \frac{K!}{(K+n-1)!}.
\end{aligned}$$

Thus, all these constants a_ψ are the same. As a result:

$$\mathbf{1} \;=\; \sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})} \frac{K!}{(K+n-1)!} \cdot \mathfrak{d}_\varphi \;=\; \frac{K!}{(K+n-1)!} \cdot \sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})} \mathfrak{d}_\varphi.$$

By moving the fraction to the other side we are done. ◀

7 Dual Dirichlet and signed hypergeometric

In the previous section we have introduced the dual basis vectors $\mathfrak{d}_\varphi$ as duals to the multinomial vectors $\mathfrak{m}_\varphi$ and have seen that each of these $\mathfrak{d}_\varphi$ forms a signed probability density. We can now start harvesting results, first by defining the associated continuous probability measure.

Definition 12 *Let $\upsilon \in \mathscr{M}(\boldsymbol{n})$ be an multiset (thought of as an urn).*

1. *We write $DDir(\upsilon)$ for the signed probability measure on Δ^n given by the density $\mathfrak{d}_\upsilon$. We call it the* dual Dirichlet *distribution.*
2. *For each number K we define the* signed hypergeometric *channel as the composite with multinomial draws*
$$shg[K] \;\coloneqq\; mn[K] \circ DDir\colon\; \mathscr{M}(\boldsymbol{n}) \to \mathscr{S}\big(\mathscr{M}[K](\boldsymbol{n})\big).$$
This means:
$$shg[K](\upsilon) \;=\; mn[K] \mathrel{\gg\!=} DDir(\upsilon) \;=\; \sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})} \left(\int_{\boldsymbol{r}\in\Delta^n} \mathfrak{m}_\varphi(\boldsymbol{r}) \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r}\right) \big|\, \varphi \,\big\rangle.$$

In Example 20 in Appendix A a signed hypergeometric distribution is computed concretely. In the multivariate case we do not have an explicit formula, as exists in the bivariate case, see Equation (13) in Appendix C for details.

The next result shows that when there is no overdrawing, there is no difference between signed and ordinary hypergeometric probabilities. This means that the dual Dirichlet distribution confirms the question that we originally set ourselves: there is a distribution over which multinomials yield hypergeometric distributions, in analogy with Theorem 5 (1).

Theorem 13 *Let urn $\upsilon \in \mathscr{M}(\boldsymbol{n})$ have size $L = \|\upsilon\|$. Then $shg[K](\upsilon) = hg[K](\upsilon)$, for each $K \le L$.*

This says that when the size of the draw is at most the size of the urn — so when there are no overdraws — signed hypergeometric coincides with ordinary hypergeometric. In particular, in this case no negative probabilities appear in the signed hypergeometric.

Proof We first note that:

$$1 \overset{(9)}{=} \big\langle \mathfrak{m}_\upsilon, \mathfrak{d}_\upsilon \big\rangle = \int_{\boldsymbol{r}\in\Delta^n} \mathfrak{m}_\upsilon(\boldsymbol{r}) \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \overset{(7)}{=} \Big(mn[L] \gg\!\!= DDir(\upsilon) \Big)(\upsilon).$$

As a result: $mn[L] \gg\!\!= DDir(\upsilon) = 1|\,\upsilon\,\rangle$. By combining this fact with Proposition 3 (3) we are done:

$$\begin{aligned} hg[K](\upsilon) \;&=\; hg[K] \gg\!\!= 1|\,\upsilon\,\rangle \;=\; hg[K] \gg\!\!= \Big(mn[L] \gg\!\!= DDir(\upsilon) \Big) \\ &=\; \Big(hg[K] \odot mn[L] \Big) \gg\!\!= DDir(\upsilon) \;=\; mn[K] \gg\!\!= DDir(\upsilon). \end{aligned}$$ ◀

Corollary 14 *Let $\upsilon \in \mathscr{M}[L](\boldsymbol{n})$ be a multiset/urn of size $L \geq 1$.*

1. *For each multiset $\varphi \leq_K \upsilon$, that is, for each $\varphi \in \mathscr{M}[K](\boldsymbol{n})$ with $\varphi \leq \upsilon$, and thus $K \leq L$,*

$$\int_{\boldsymbol{r}\in\Delta^n} \boldsymbol{r}^{\varphi} \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \;=\; \frac{\upsilon! \cdot (L-K)!}{(\upsilon-\varphi)! \cdot L!}.$$

2. *In particular, for each $i \in supp(\upsilon) \subseteq \boldsymbol{n}$.*

$$\int_{\boldsymbol{r}\in\Delta^n} r_i \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \;=\; flrn(\upsilon)(i).$$

Proof 1. Since:

$$\begin{aligned} \int_{\boldsymbol{r}\in\Delta^n} \boldsymbol{r}^{\varphi} \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \;&=\; \frac{1}{\big(\varphi\big)} \cdot \int_{\boldsymbol{r}\in\Delta^n} \mathfrak{m}_\varphi(\boldsymbol{r}) \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \;=\; \frac{1}{\big(\varphi\big)} \cdot \Big(mn[K] \gg\!\!= DDir(\upsilon) \Big)(\varphi) \\ &=\; \frac{1}{\big(\varphi\big)} \cdot hg[K](\upsilon)(\varphi) \;=\; \frac{\varphi!}{K!} \cdot \frac{\binom{\upsilon}{\varphi}}{\binom{L}{K}} \;=\; \frac{\upsilon! \cdot (L-K)!}{(\upsilon-\varphi)! \cdot L!}. \end{aligned}$$

2. We apply the previous point with $\varphi = 1|\,i\,\rangle$ of size 1. Then:

$$\int_{\boldsymbol{r}\in\Delta^n} r_i \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \;=\; \int_{\boldsymbol{r}\in\Delta^n} \boldsymbol{r}^{1|\,i\,\rangle} \cdot \mathfrak{d}_\upsilon(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} \;=\; \frac{\upsilon! \cdot (L-1)!}{(\upsilon - 1|\,i\,\rangle)! \cdot L!} \;=\; \frac{\upsilon(i)}{L} \;=\; \frac{\upsilon(i)}{\|\upsilon\|} \;=\; flrn(\upsilon)(i).$$ ◀

Proposition 15 1. *We write* **0** *for the empty multiset and* **1** *for the multiset of singletons, say on $\boldsymbol{n} = \{0,\ldots,n-1\}$. Then $DDir(\mathbf{0}) = Dir(\mathbf{1})$ is the uniform measure on Δ^n.*

2. *The sample channel* $\mathrm{sam}\colon \Delta^n \to \mathscr{D}(\boldsymbol{n})$ *from Theorem 5* (4) *gives, like for ordinary Dirichlet,*

$$sam \gg\!\!= DDir(\upsilon) \;=\; flrn(\upsilon).$$

Proof 1. For $K = 0$ the set $\mathscr{M}[K](\boldsymbol{n})$ of multisets of size 0 contains the empty multiset **0** as sole element. The associated factor sum matrix FS from (12) is thus the singleton matrix (1), with (1) as inverse. Hence the only coefficient $c_{\mathbf{0},\mathbf{0}}$ of the polynomial $\mathfrak{d}_{\mathbf{0}}$ in (12) is $(n-1)!$. This makes $\mathfrak{d}_{\mathbf{0}}$ the constant function $\boldsymbol{r} \mapsto (n-1)!$, which is the density of the uniform measure $Dir(\mathbf{1})$ on Δ^n.

2. Using Corollary 14 (2), the reasoning is precisely as in the proof of Theorem 5 (4). ◀

Theorem 13 says that the signed hypergeometric distribution is a 'conservative' extension of the ordinary hypergeometric distribution in the sense that these distributions coincide for draws of size below the size of the urn. We now illustrate draws of arbitrary size, also bigger than the size of the urn. The physical interpretation of such overdraws is unclear. But mathematically all works well.

We continue with some basic properties of signed hypergeometric distributions, as analogues of (some of the items of) Proposition 3.

Proposition 16 *1. $shg[K](\mathbf{0})$ is the uniform distribution on $\mathcal{M}[K](\boldsymbol{n})$;*

2. *$flrn \circ shg[K] = flrn$;*
3. *$shg[K] \circ mn[L{+}K] = mn[K]$;*
4. *$shg[K] \circ shg[K{+}L] = shg[K]$;*
5. *$DD \circ shg[K{+}1] = shg[K]$.*

The last equation shows that the signed hypergeometric form a cone for the draw-delete maps and thus fit in a categorial approach to 'De Finetti', following [15]. The equation $hg[K] \circ DD = hg[K]$ from Proposition 3 (5) holds for ordinary hypergeometric, but its analogue for signed hypergeometric fails.

Proof 1. Via Proposition 15 (1): $shg[K](\mathbf{0}) = mn[K] \gg DDir(\mathbf{0}) = mn[K] \gg Dir(\mathbf{1}) = pol[K](\mathbf{1})$. The latter Pólya distribution on $\mathcal{M}[K](\boldsymbol{n})$ is uniform, see Theorem 5.

2. By combining Proposition 3 (2) with Proposition 15 (2) we get: $flrn \circ shg[K] = flrn \circ mn[K] \circ DDir = sam \circ DDir = flrn$.

3. In the composite $shg[K] \circ mn[L{+}K]$ the signed hypergeometric draws of size K are applied to the urns that appear as draws of size $L{+}K$ coming out of the multinomial $mn[L{+}K]$. Hence the draw size is less than the urn size, so Theorem 13 applies, and the signed hypergeometric is an ordinary hypergeometric. Thus Proposition 3 (3) gives: $shg[K] \circ mn[L{+}K] = hg[K] \circ mn[L{+}K] = mn[K]$.

4. By the previous point: $shg[K] \circ shg[L{+}K] = shg[K] \circ mn[L{+}K] \circ DDir = mn[K] \circ DDir = shg[K]$.

5. Via Proposition 3 (7): $DD \circ shg[K{+}1] = DD \circ mn[K{+}1] \circ DDir = mn[K] \circ DDir = shg[K]$. ◀

8 Signed hypergeometric channels as Bayesian inversion

This section shows that the signed hypergeometric channel $shg[K]\colon \mathcal{M}(\boldsymbol{n}) \to \mathcal{S}\big(\mathcal{M}[K](\boldsymbol{n})\big)$ can be obtained as dagger, that is as Bayesian inversion, see [4, 3]. For this we need the following new signed distribution, that builds on the result from Proposition 11 that the sum of dual basis functions is constant.

Definition 17 *For a distribution $\omega \in \mathcal{D}(\boldsymbol{n})$ we define signed* dual multinomial *distribution* $dmn[K](\omega) \in \mathcal{S}\big(\mathcal{M}[K](\boldsymbol{n})\big)$ *as:*

$$dmn[K](\omega) := \sum_{\varphi \in \mathcal{M}[K](\boldsymbol{n})} \frac{K!}{(K{+}n{-}1)!} \cdot \mathfrak{d}_{\varphi}(\omega)\,\big|\,\varphi\,\big\rangle.$$

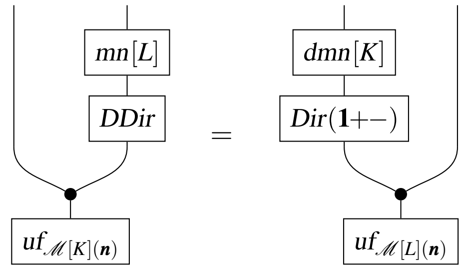
We thus get a 'signed' channel $dmn[K]\colon \mathcal{D}(\boldsymbol{n}) \to \mathcal{S}\big(\mathcal{M}(\boldsymbol{n})\big)$.

For instance:

$$\begin{aligned} & dmn[2]\Big(\tfrac{1}{2}|0\rangle + \tfrac{1}{6}|1\rangle + \tfrac{1}{3}|2\rangle\Big) \\ &= -\tfrac{1}{4}\Big|\,2|0\rangle\Big\rangle + \tfrac{1}{6}\Big|\,1|0\rangle + 1|1\rangle\Big\rangle - \tfrac{1}{4}\Big|\,2|1\rangle\Big\rangle + \tfrac{11}{6}\Big|\,1|0\rangle + 1|2\rangle\Big\rangle + \tfrac{1}{6}\Big|\,1|1\rangle + 1|2\rangle\Big\rangle - \tfrac{2}{3}\Big|\,2|2\rangle\Big\rangle\,. \end{aligned}$$

We have no (operational) interpretation for these distributions, but they do make sense mathematically, as in the next result.

Theorem 18 *1. There is an equality of string diagrams which involves swapping dual and ordinary Dirichlet distributions*



2. *The signed hypergeometric channel* $shg[L]\colon \mathscr{M}[K](\boldsymbol{n}) \to \mathscr{S}\big(\mathscr{M}[L](\boldsymbol{n})\big)$ *is the dagger of the composite* $dmn[K] \odot Dir(\mathbf{1}{+}{-})\colon \mathscr{M}[L](\boldsymbol{n}) \to \mathscr{S}\big(\mathscr{M}[K](\boldsymbol{n})\big)$ *with the uniform distribution* $\mathrm{uf}_{\mathscr{M}[L](\boldsymbol{n})}$ *as prior. In a formula:*

$$shg[L] \;=\; \Big(dmn[K] \odot Dir(\mathbf{1}{+}{-})\Big)^{\dagger}_{\mathrm{uf}_{\mathscr{M}[L](\boldsymbol{n})}}.$$

Proof 1. Let $\varphi \in \mathscr{M}[K](\boldsymbol{n})$ and $\psi \in \mathscr{M}[L](\boldsymbol{n})$.

$$\begin{aligned}
\Big(\langle dmn[K] \odot Dir(\mathbf{1}{+}{-}), id\rangle \gg\!= \mathrm{uf}_{\mathscr{M}[L](\boldsymbol{n})}\Big)(\varphi,\psi) &= \frac{1}{\left(\!\binom{n}{L}\!\right)} \cdot \int_{\boldsymbol{r}} \frac{K!}{(K{+}n{-}1)!} \cdot \mathfrak{d}_{\varphi}(\omega) \cdot \frac{(L{+}n{-}1)!}{\psi\,!!} \cdot \boldsymbol{r}^{\psi}\,\mathrm{d}\boldsymbol{r} \\
&= \frac{K! \cdot (n{-}1)!}{(K{+}n{-}1)!} \cdot \int_{\boldsymbol{r}} \mathfrak{d}_{\varphi}(\omega) \cdot \frac{L!}{\psi\,!!} \cdot \boldsymbol{r}^{\psi}\,\mathrm{d}\boldsymbol{r} \\
&= \frac{1}{\left(\!\binom{n}{K}\!\right)} \cdot \int_{\boldsymbol{r}} \mathfrak{d}_{\varphi}(\boldsymbol{r}) \cdot \mathfrak{m}_{\psi}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \\
&= \frac{1}{\left(\!\binom{n}{K}\!\right)} \cdot \Big(mn[L] \gg\!= DDir(\varphi)\Big)(\psi) \\
&= \Big(\langle id\,, mn[L] \odot DDir\rangle \gg\!= \mathrm{uf}_{\mathscr{M}[K](\boldsymbol{n})}\Big)(\varphi,\psi).
\end{aligned}$$

2. This is a reformulation of the previous point, using that $shg[L] = mn[L] \odot DDir$, occurring in the above string diagram on the left of the equation. ◀

In Definition 12 we have introduced the signed hypergeometric $shg[L]$ as $mn[L] \odot DDir$ via the dual Dirichlet distribution $DDir$. The above result tells that we can also obtain $shg[L]$ as dagger from ordinary Dirichlet Dir (plus the dual multinomial distribution dmn). This diagrammatic description of the dagger coincides with the one on right in Figure 1.

9 Conclusions and further work

This paper covers a fascinating topic, namely negative probabilities. It does not offer operational meaning, but it does provide a solid mathematical basis for the emergence of negative probabilities in classical, non-quantum probability theory. The techniques of categorical probability provide the toolbox for describing the relevant properties.

There is plenty of further work. High on our list is an explicit formula for the dual basis vectors in the general multivariate case (like in the bivariate case). Also, we would like to develop a deeper understanding of the conjugate situation described in the string diagrams in Theorem 18.

References

[1] S. Abramsky & A. Brandenburger (2014): *An Operational Interpretation of Negative Probabilities and No-Signalling Models*. In F. van Breugel, E. Kashefi, C. Palamidessi & J. Rutten, editors: *Horizons of the Mind. A Tribute to Prakash Panangaden*, *Lect. Notes Comp. Sci.* 8464, Springer, Berlin, pp. 59–75, doi:10.1007/978-3-319-06880-0_3.

[2] A. Blass & Y. Gurevich (2021): *Negative probabilities: what are they for?* *Journal of Physics A: Mathematical and Theoretical* 54(31), p. 315303.

[3] K. Cho & B. Jacobs (2019): *Disintegration and Bayesian inversion via string diagrams*. *Math. Struct. in Comp. Sci.* 29(7), pp. 938–971, doi:10.1017/s0960129518000488.

[4] F. Clerc, F. Dahlqvist, V. Danos & I. Garnier (2017): *Pointless learning*. In J. Esparza & A. Murawski, editors: *Foundations of Software Science and Computation Structures*, *Lect. Notes Comp. Sci.* 10203, Springer, Berlin, pp. 355–369, doi:10.1007/978-3-662-54458-7_21.

[5] P. Diaconis (1977): *Finite forms of De Finetti's theorem on exchangeability*. *Synthese* 36(2), pp. 271–281.

[6] W. Dong-Bing (1993): *Dual bases of a Bernstein polynomial basis on simplices*. *Computer aided geometric design* 10(6), pp. 483–489.

[7] R. Feynman (1987): *Negative Probability*. In B. Hiley & F. Peat, editors: *Quantum Implications, Essays in Honor of David Bohm*, Routledge and Kegan Paul, London, pp. 235–246.

[8] T. Fritz (2020): *A synthetic approach to Markov kernels, conditional independence, and theorems on sufficient statistics*. *Advances in Math.* 370, p. 107239, doi:10.1016/J.AIM.2020.107239.

[9] J. Hoschek & D. Lasser (1993): *Fundamentals of computer aided geometric design*. AK Peters, Ltd.

[10] B. Jacobs (2016): *Affine Monads and Side-Effect-Freeness*. In I. Hasuo, editor: *Coalgebraic Methods in Computer Science (CMCS 2016)*, *Lect. Notes Comp. Sci.* 9608, Springer, Berlin, pp. 53–72, doi:10.1007/978-3-319-40370-0_5.

[11] B. Jacobs (2018): *From Probability Monads to Commutative Effectuses*. *Journ. of Logical and Algebraic Methods in Programming* 94, pp. 200–237, doi:10.1016/j.jlamp.2016.11.006.

[12] B. Jacobs (2020): *A Channel-Based Perspective on Conjugate Priors*. *Math. Struct. in Comp. Sci.* 30(1), pp. 44–61, doi:10.1017/S0960129519000082.

[13] B. Jacobs (2021): *From Multisets over Distributions to Distributions over Multisets*. In: *Logic in Computer Science*, IEEE, Computer Science Press, doi:10.1109/lics52264.2021.9470678.

[14] B. Jacobs (2022): *Urns & Tubes*. *Compositionality* 4(4), doi:10.32408/compositionality-4-4.

[15] B. Jacobs & S. Staton (2020): *De Finetti's construction as a categorical limit*. In D. Petrişan & J. Rot, editors: *Coalgebraic Methods in Computer Science (CMCS 2020)*, *Lect. Notes Comp. Sci.* 12094, Springer, Berlin, pp. 90–111, doi:10.1007/978-3-030-57201-3_6.

[16] E. Jaynes (1982): *Some applications and extensions of the De Finetti representation theorem*. *Bayesian Inference and Decision Techniques with Applications: Essays in Honor of Bruno de Finetti. North-Holland Publishers.*

[17] B. Jüttler (1998): *The dual basis functions for the Bernstein polynomials*. *Adv. Computational Mathematics* 8, pp. 345–352, doi:10.1023/A:1018912801267.

[18] G. Kerns & G. Székely (2006): *De Finetti's Theorem for abstract finite exchangeable sequences*. *Journal of Theoretical Probability* 19(3), pp. 589–608.

[19] G. Lorentz (2013): *Bernstein polynomials*. American Mathematical Soc.

[20] G. Meissner & M. Burgin (2011): *Negative Probabilities in Financial Modeling*. *Wilmott Magazine*, doi:10.2139/ssrn.1773077.

[21] P. Panangaden (2009): *Labelled Markov Processes*. Imperial College Press, London.

[22] G. Székely (2005): *Half of a Coin: Negative Probabilities*. *Wilmott Magazine*, pp. 66–68.

[23] H. Tijms & K. Staats (2007): *Negative probabilities at work in the M/D/1 queue*. Probability in the Engineering and Informational Sciences 21(1), pp. 67–76.

[24] K. Zhao & J. Sun (1988): *Dual bases of multivariate Bernstein-Bézier polynomials*. Computer aided geometric design 5(2), pp. 119–125.

Appendix

A Examples

We elaborate two examples: first we illustrate how to actually compute a dual basis, and then how to compute signed hypergeometric distributions.

Example 19 *We take $n=3$ and $K=2$ so that we have $\left(\!\binom{n}{K}\!\right) = \frac{4!}{2!\cdot 2!} = 6$ multisets in $\mathcal{M}[2](\mathbf{3})$, namely: $2|0\rangle,\ 1|0\rangle+1|1\rangle,\ 2|1\rangle,\ 1|0\rangle+1|2\rangle,\ 1|1\rangle+1|2\rangle,\ 2|2\rangle$. Using this order of multisets we get a 6×6 matrix:*

$$FS \;=\; \Big((\varphi+\psi)\,!!\Big)_{\varphi,\psi\in\mathcal{M}[2](\mathbf{3})} \;=\; \begin{pmatrix} 24 & 6 & 4 & 6 & 2 & 4 \\ 6 & 4 & 6 & 2 & 2 & 2 \\ 4 & 6 & 24 & 2 & 6 & 4 \\ 6 & 2 & 2 & 4 & 2 & 6 \\ 2 & 2 & 6 & 2 & 4 & 6 \\ 4 & 2 & 4 & 6 & 6 & 24 \end{pmatrix} \qquad so \qquad FS^{-1} \;=\; \frac{1}{60}\begin{pmatrix} 6 & -8 & 1 & -8 & 2 & 1 \\ -8 & 44 & -8 & -6 & -6 & 2 \\ 1 & -8 & 6 & 2 & -8 & 1 \\ -8 & -6 & 2 & 44 & -6 & -8 \\ 2 & -6 & -8 & -6 & 44 & -8 \\ 1 & 2 & 1 & -8 & -8 & 6 \end{pmatrix}$$

Notice that negative values appear in this inverse matrix, without clear pattern.

We can now compute the dual basis vectors by combining (10) *and* (12). *We elaborate the case of the first multiset $2|0\rangle$. For $(r_0,r_1,r_2)\in\Delta^3$, that is, for $r_0,r_1,r_2\in[0,1]$ with $r_0+r_1+r_2=1$ we get $\mathfrak{d}_{2|0\rangle}\in P_2(\Delta^3)$ determined as:*

$$\begin{aligned}
&\mathfrak{d}_{2|0\rangle}(r_0,r_1,r_2) \\
&\overset{(10)}{=} \sum_{\varphi\in\mathcal{M}[2](\mathbf{3})} c_{\varphi,2|0\rangle}\cdot(r_0,r_1,r_2)^{\varphi} \overset{(12)}{=} \sum_{\varphi\in\mathcal{M}[2](\mathbf{3})} \tfrac{6!}{(2|0\rangle)}\cdot\Big(FS^{-1}\Big)_{\varphi,2|0\rangle}\cdot r_0^{\varphi(0)}\cdot r_1^{\varphi(1)}\cdot r_2^{\varphi(2)} \\
&= \; 12\Big(6r_0^2-8r_0^1r_1^1+1r_1^2-8r_0^1r_2^1+2r_1^1r_2^1+1r_2^2\Big) \;=\; 72r_0^2-96r_0r_1+12r_1^2-96r_0r_2+24r_1r_2+12r_2^2.
\end{aligned}$$

Similarly, for the other multisets in $\mathcal{M}[2](\mathbf{3})$,

$$\begin{aligned}
\mathfrak{d}_{1|0\rangle+1|1\rangle}(r_0,r_1,r_2) &= -48r_0^2+264r_0r_1-48r_1^2-36r_0r_2-36r_1r_2+12r_2^2 \\
\mathfrak{d}_{2|1\rangle}(r_0,r_1,r_2) &= 12r_0^2-96r_0r_1+72r_1^2+24r_0r_2-96r_1r_2+12r_2^2 \\
\mathfrak{d}_{1|0\rangle+1|2\rangle}(r_0,r_1,r_2) &= -48r_0^2-36r_0r_1+12r_1^2+264r_0r_2-36r_1r_2-48r_2^2 \\
\mathfrak{d}_{1|1\rangle+1|2\rangle}(r_0,r_1,r_2) &= 12r_0^2-36r_0r_1-48r_1^2-36r_0r_2+264r_1r_2-48r_2^2 \\
\mathfrak{d}_{2|2\rangle}(r_0,r_1,r_2) &= 12r_0^2+24r_0r_1+12r_1^2-96r_0r_2-96r_1r_2+72r_2^2.
\end{aligned}$$

Then indeed, $\big\langle \mathfrak{m}_{\varphi},\mathfrak{d}_{\psi}\big\rangle = \delta_{\varphi,\psi}$ for all $\varphi,\psi\in\mathcal{M}[2](\mathbf{3})$.

We check the claim of Proposition 11, that the sum of the dual base vectors is a particular constant. Let's abbreviate the (pointwise) sum of the above dual basis functions as: $\eth := \sum_{\varphi\in\mathcal{M}[2](\mathbf{3})} \eth_\varphi$. *Then, using the above descriptions, for* $(r_0,r_1,r_2)\in\Delta^3$,

$$\begin{aligned}\eth(r_0,r_1,r_2) &= 12r_0^2+24r_0r_1+12r_1^2+24r_0r_2+24r_1r_2+12r_2^2 \\ &= 12\Big((r_0+r_1)^2+2(r_0+r_1)r_2+r_2^2\Big) = 12\big(r_0+r_1+r_2\big)^2 = 12.\end{aligned}$$

Hence the sum of the dual basis functions is indeed a constant and equals $\frac{(K+n-1)!}{K!} = \frac{4!}{2!} = 12$.

Example 20 *Let's take as urn* $\upsilon = 1|0\rangle + 1|1\rangle + 1|2\rangle$ *with one ball of each color in* $\mathbf{3} = \{0,1,2\}$. *By drawing two balls we remain within the world of ordinary hypergeometric distributions, by Theorem 13 and get the following distribution over draws.*

$$shg[2](\upsilon) = hg[2](\upsilon) = \tfrac{1}{3}\big|1|0\rangle+1|1\rangle\big\rangle + \tfrac{1}{3}\big|1|0\rangle+1|2\rangle\big\rangle + \tfrac{1}{3}\big|1|1\rangle+1|2\rangle\big\rangle.$$

We can also draw three balls; in that case, the outcome is certain to be the whole urn:

$$shg[3](\upsilon) = hg[3](\upsilon) = 1\big|1|0\rangle+1|1\rangle+2|1\rangle\big\rangle = 1\big|\upsilon\big\rangle.$$

Drawing four balls, more than in the urn, is only possible with the signed hypergeometric. It leads to negative probabilities in:

$$\begin{aligned}shg[4](\upsilon) = {} & \tfrac{7}{126}\big|4|0\rangle\big\rangle - \tfrac{14}{126}\big|3|0\rangle+1|1\rangle\big\rangle - \tfrac{21}{126}\big|2|0\rangle+2|1\rangle\big\rangle - \tfrac{14}{126}\big|1|0\rangle+3|1\rangle\big\rangle + \tfrac{7}{126}\big|4|1\rangle\big\rangle \\ & - \tfrac{14}{126}\big|3|0\rangle+1|2\rangle\big\rangle + \tfrac{84}{126}\big|2|0\rangle+1|1\rangle+1|2\rangle\big\rangle + \tfrac{84}{126}\big|1|0\rangle+2|1\rangle+1|2\rangle\big\rangle \\ & - \tfrac{14}{126}\big|3|1\rangle+1|2\rangle\big\rangle - \tfrac{21}{126}\big|2|0\rangle+2|2\rangle\big\rangle + \tfrac{84}{126}\big|1|0\rangle+1|1\rangle+2|2\rangle\big\rangle \\ & - \tfrac{21}{126}\big|2|1\rangle+2|2\rangle\big\rangle - \tfrac{14}{126}\big|1|0\rangle+3|2\rangle\big\rangle - \tfrac{14}{126}\big|1|1\rangle+3|2\rangle\big\rangle + \tfrac{7}{126}\big|4|2\rangle\big\rangle.\end{aligned}$$

There is no apparent 'logic' in these probabilities, for instance in terms of draw probabilities. In particular, an intuitive explanation is missing for why certain probabilities are negative. But, as explained above, these distributions are constructed in a systematic manner, via dual bases.

Clarification: the outcomes of the signed hypergeometric given above are obtained via elementary Python scripts that perform matrix inversion — as in (12) *— to obtain representations of dual basis vectors and to perform integration over simplices. Our Python scripts produce real numbers as probabilities, but they are so close to the above fractions in* $shg[4](\upsilon)$ *that we write these fractions instead, for the sake of readability.*

B Missing proofs

In the body of the article we left out the proofs of several basis properties of ordinary draws.

Proof (of Theorem 5)

1. Via (7), for arbitrary $\varphi \in \mathscr{M}[K](\boldsymbol{n})$,

$$\begin{aligned}
\big(mn[K] \gg\!\!= Dir(\upsilon)\big)(\varphi) &= \int_{\boldsymbol{r}\in\Delta^n} mn[K](\boldsymbol{r})(\varphi)\cdot dir(\upsilon)(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \\
&= \int_{\boldsymbol{r}\in\Delta^n} (\varphi)\cdot\prod_{i\in\boldsymbol{n}} r_i^{\varphi(i)}\cdot\frac{(L-1)!}{(\upsilon-\mathbf{1})\underset{\circ}{!}}\cdot\prod_{i\in\boldsymbol{n}} r_i^{\upsilon(i)-1}\,\mathrm{d}\boldsymbol{r} \\
&= \int_{\boldsymbol{r}\in\Delta^n} \frac{K!\cdot(L-1)!}{\varphi\underset{\circ}{!}\cdot(\upsilon-\mathbf{1})\underset{\circ}{!}}\cdot\prod_{i\in\boldsymbol{n}} r_i^{\upsilon(i)+\varphi(i)-1}\,\mathrm{d}\boldsymbol{r} \\
&= \frac{\big(\!\binom{\upsilon}{\varphi}\!\big)}{\big(\!\binom{L}{K}\!\big)}\cdot\int_{\boldsymbol{r}\in\Delta^n}\frac{(L+K-1)!}{(\upsilon+\varphi-\mathbf{1})\underset{\circ}{!}}\cdot\prod_{i\in\boldsymbol{n}} r_i^{(\upsilon+\varphi)(i)-1}\,\mathrm{d}\boldsymbol{r} \\
&\overset{(6)}{=} pol[K](\upsilon)(\varphi).
\end{aligned}$$

2. This can be computed in a similar manner, but the details are beyond the scope of this paper.

3. We have already seen that the set $\mathscr{M}[K](\boldsymbol{n})$ has $\big(\!\binom{n}{K}\!\big)$ elements, so that the uniform distribution $uf_{\mathscr{M}[K](\boldsymbol{n})}$ on this set is $\sum_{\varphi\in\mathscr{M}[K](\boldsymbol{n})}\frac{1}{(\binom{n}{K})}\big|\,\varphi\big\rangle$. This uniform distribution equals $pol[K](\mathbf{1})$, where $\mathbf{1} = \sum_{i\in\boldsymbol{n}} 1|\,i\rangle$. Hence we can reason diagrammatically, as on the right in Figure 1. The uniform distribution $uf_{\mathscr{D}(\boldsymbol{n})}$ on the left in this chain of equations is $Dir(\mathbf{1})$ on Δ^n, with density $\boldsymbol{r}\mapsto(n-1)!$, since $\int_{\boldsymbol{r}\in\Delta^n} 1\,\mathrm{d}\boldsymbol{r} = \frac{1}{(n-1)!}$ by (6).

4. Consider an urn υ of size L and a number $j\in\boldsymbol{n}$. We write $\upsilon_j := \upsilon + 1|\,j\,\rangle$ of size $L+1$. Then, using (7),

$$\begin{aligned}
\big(sam \gg\!\!= Dir(\upsilon)\big)(j) &= \int_{\boldsymbol{r}\in\Delta^n} r_j\cdot\frac{(L-1)!}{(\upsilon-\mathbf{1})\underset{\circ}{!}}\cdot\prod_{i\in\boldsymbol{n}} r_i^{\upsilon(i)-1}\,\mathrm{d}\boldsymbol{r} \\
&= \frac{\upsilon(j)}{L}\cdot\int_{\boldsymbol{r}\in\Delta^n}\frac{L!}{(\upsilon_j-\mathbf{1})\underset{\circ}{!}}\cdot\prod_{i\in\boldsymbol{n}} r_i^{\upsilon_j(i)-1}\,\mathrm{d}\boldsymbol{r} \\
&\overset{(6)}{=} \frac{\upsilon(j)}{L}\cdot 1 \\
&= flrn(\upsilon).
\end{aligned}$$
◀

C The bivariate case

In the previous sections we have elaborated the multivariate case, involving multiple variables. The bivariate case, for $n = 2$, can be done slightly differently, using the isomorphisms $\mathscr{D}(\mathbf{2}) \cong [0,1]$ and $\mathscr{M}[K](\mathbf{2}) \cong \{0,1,\ldots,K\} \cong \boldsymbol{K+1}$. The binomial channel $bn[K]\colon [0,1] \rightsquigarrow \{0,\ldots,K\}$ is thus related via the general multinomial one via the following square.

$$\begin{tikzcd}
{[0,1]} \arrow[r, "bn{[K]}"] \arrow[d, "\cong"] & \{0,\ldots,K\} \\
\mathscr{D}(\mathbf{2}) \arrow[r, "mn{[K]}"] & \mathscr{M}[K](\mathbf{2}) \arrow[u, "\cong"']
\end{tikzcd}$$

Explicitly, for $r \in [0,1]$, we have

$$bn[K](r) = \sum_{0 \le i \le K} \binom{K}{i} \cdot r^i \cdot (1-r)^{K-i} \big| i \big\rangle.$$

The polynomials involved are known as *Bernstein polynomials*, namely

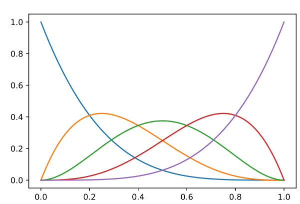
$$\begin{aligned} \mathfrak{b}_i(r) \coloneqq \binom{K}{i} \cdot r^i \cdot (1-r)^{K-i} &= \sum_{0 \le j \le K-i} \binom{K}{i} \cdot \binom{K-i}{j} \cdot r^i \cdot 1^j \cdot (-r)^{K-i-j} \\ &= \sum_{0 \le j \le K-i} \binom{K}{i} \cdot \binom{K-i}{j} \cdot (-1)^{K-i-j} \cdot r^{K-j}. \end{aligned}$$

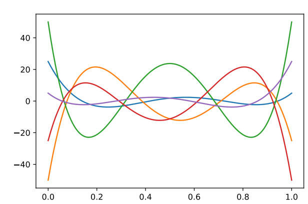
These polynomials are widely studied in Computer Graphics and Computer Aided Geometric Design [9], but also in areas such as approximation theory [19] and probability. They appear not only as probability mass functions for the binomial distributions (as described above), but also as density function of the (continuous) Beta distributions, rescaled by a normalisation factor.

The Hilbert space $P_K(\Delta^2)$ can be identified with the space of univariate polynomials, as functions $[0,1] \to \mathbb{R}$, spanned by the monomial basis $(r^i : 0 \le i \le K)$. The $(n+1) \times (n+1)$ matrix B that represents the above polynomials $\mathfrak{b}_i$ in this basis has entries

$$B_{i,K-j} = \binom{K}{i} \cdot \binom{K-i}{j} \cdot (-1)^{K-i-j}.$$

Below we plot several Bernstein polynomials (on the left) and their dual bases (on the right).





We write S for the matrix with inner products of the base vectors, so:

$$S_{i,j} \coloneqq \langle r^i, r^j \rangle = \int_{r \in [0,1]} r^i \cdot r^j \, \mathrm{d}r = \tfrac{1}{i+j+1}.$$

The dual basis of the binomial polynomials $\mathfrak{b}_i$ are then given by the matrix inverse $(B^T \cdot S)^{-1}$. It has been studied in a Computer Graphics context [6] [24], and formulas for computing the dual basis are known, see [17] for an overview.

This explicit formulation of the dual basis allows us to describe the *bivariate signed hypergeometric* channel $bshg[L,K] \colon \{0,\ldots,L\} \to \mathscr{S}\big(\{0,\ldots,K\}\big)$, in the following commuting diagram.

$$\begin{tikzcd} \{0,\ldots,L\}_{\cong} \arrow[r, "bshg{[L,K]}", "\circ" marking] \arrow[d] & \{0,\ldots,K\} \\ \mathscr{M}[L](\mathbf{2}) \arrow[r, "shg{[K]}", "\circ" marking] & \mathscr{M}[K](\mathbf{2})_{-\cong} \arrow[u] \end{tikzcd}$$

We include the parameter L in writing $bshg[L,K]$ since it cannot be derived from an input $j \in \{0,\ldots,L\}$. In contrast, writing this parameter explicitly is not needed in the multivariate case, since the size of the urn can be computed from the urn itself.

The explicit formula for this bivariate signed hypergeometric $bshg[L,K]$ is:

$$\begin{aligned} & bshg[L,K](j) \\ & := \sum_{0\leq i\leq K} \frac{\binom{K}{i}}{(K+L+1)\cdot\binom{L}{j}} \left(\sum_{0\leq \ell\leq L} \frac{(-1)^{j+\ell}}{\binom{K+L}{i+\ell}} \sum_{0\leq k\leq \min(j,\ell)} (2k+1) \binom{L+k+1}{L-j}\binom{L-k}{L-j}\binom{L+k+1}{L-\ell}\binom{L-k}{L-\ell} \right) \big| i \big\rangle \end{aligned} \tag{13}$$

This $bshg[L,K]$ is defined on $0 \leq j \leq L$, corresponding to urn $j|0\rangle + (L-j)|1\rangle$. The complicated character of this formula is not helpful for an operational interpretation in terms of draw probabilities. But it does allow us to compute (truely) exact distributions.

Example 21 *Let's take an urn size $L = 3$ with 2 balls of colour 0. Thus, in a multivariate scenario we would write this as urn $\upsilon = 2|0\rangle + 1|1\rangle$. We first look at a draw of size $K = 4$. Using the formula from Figure C we get a bivariate signed hypergeometric distribution of the form:*

$$bshg[3,4](2) \;=\; \tfrac{17}{210}\big|0\big\rangle - \tfrac{34}{105}\big|1\big\rangle + \tfrac{17}{35}\big|2\big\rangle + \tfrac{106}{105}\big|3\big\rangle - \tfrac{53}{210}\big|4\big\rangle.$$

The number i in $|i\rangle$ refers to the number of balls of colour 0 that are drawn (out of K in total), with corresponding (positive or negative) probability.

Similarly, for $K = 5$ we have:

$$bshg[3,5](2) \;=\; \tfrac{1}{6}\big|0\big\rangle - \tfrac{3}{7}\big|1\big\rangle + \tfrac{1}{21}\big|2\big\rangle + \tfrac{16}{21}\big|3\big\rangle + \tfrac{37}{42}\big|4\big\rangle - \tfrac{3}{7}\big|5\big\rangle.$$