Stick Breaking, in Coalgebra and Probability*

Bart Jacobs

Institute for Computing and Information Sciences,
Radboud University, Nijmegen, NL
bart@cs.ru.nl
March 22, 2022

Abstract. Stick breaking is an elementary operation that has been formulated and used within stochastic process theory. This paper extracts the essentials of stick breaking in terms of isomorphisms between discrete probability distributions (with full support) and sequences of numbers between zero and one. This works for both finite and infinite distributions. Stick breaking is a repetitive construction with a strong coalgebraic flavour. Indeed, it is shown that stick breaking turns discrete distributions with infinite full support on the natural numbers into a final coalgebra. Once isolated as a separate construction, the usefulness of stick breaking is illustrated in the description of various probability distributions, such as binomial & multinomial and beta & Dirichlet.

1 Introduction

Consider the following mixture of paints, of four different colours: a quarter of red (R), a third of green (G), also a quarter of blue (B) and finally a sixth of yellow (Y). We write this 'convex' combination as:

$$\tfrac{1}{4}|R\rangle + \tfrac{1}{3}|G\rangle + \tfrac{1}{4}|B\rangle + \tfrac{1}{6}|Y\rangle.$$

The ket notation $|-\rangle$ is meaningless syntactic sugar, used to separate the fractions from the colours. This combination is called 'convex' since the probabilities add up to one. We call this convex combination a (discrete, finite) probability distribution over the set of colours $\{R, G, B, Y\}$. Let's write $\mathcal{D}_{fs}\big(\{R, G, B, Y\}\big)$ for the set of all such distributions:

$$\mathcal{D}_{fs}\big(\{R, G, B, Y\}\big) = \Big\{ r_0|R\rangle + r_1|G\rangle + r_2|B\rangle + r_3|Y\rangle \;\Big|\; r_0, r_1, r_2, r_3 \in (0,1) \text{ with } r_0 + r_1 + r_2 + r_3 = 1 \Big\}.$$

We use the subscript *fs*, for 'full support'; this means that none of the r_i may be zero. It is needed below to prevent division by zero. We enforce fullness of support by requiring that the r_i are in the open unit interval $(0, 1) \subseteq \mathbb{R}$, without endpoints.

* Pre-proceedings version for CMCS 2022.

The above equation describes the set of distributions (on these four colours) as a *simplex*, of dimension *three*. Indeed, it is easy to see that one of the r_i is superfluous, since it is determined by the others. Explicitly, there is an isomorphism:

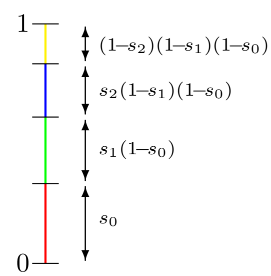
$$\mathcal{D}_{\mathit{fs}}\big(\{R, G, B, Y\}\big) \;\cong\; \Big\{(r_0, r_1, r_2) \in (0,1)^3 \;\Big|\; r_0 + r_1 + r_2 < 1\Big\}.$$

The above set on the right-hand-side is clearly a proper subset of the cube $(0,1)^3$. In essence, the *stick breaking* construction that plays a central role in this paper provides an isomorphism:

$$\mathcal{D}_{\mathit{fs}}\big(\{R, G, B, Y\}\big) \;\cong\; (0,1)^3. \tag{1}$$

This may not be immediate at first sight. One has to do (appropriate) rescaling.

There is an intuitive explanation of stick breaking in terms of successively breaking up a stick. We adapt this account to the above set of four colours. We start from three numbers $s_0, s_1, s_2 \in (0,1)$ and intend to turn them into a distribution on the set of colour $\{R, G, B, Y\}$.



Imagine a stick of length one, as described vertically on the right. We take our first number $s_0 \in (0,1)$ and decide to paint the lower part/proportion s_0 red. We now have an unpainted part of length $1 - s_0$. We paint the s_1 proportion of it green. The newly painted part then has length $s_1(1 - s_0)$. The unpainted part is now $(1 - s_2)(1 - s_0)$. We paint the s_2-proportion of this remainder blue. The final remainder is then of length $(1 - s_2)(1 - s_2)(1 - s_0)$. We paint it yellow. Note that the resulting distribution has full support.

This construction can also be described in terms of breaking a stick, at each position where we have a change of colour in the above picture. The effect is a map $(0,1)^3 \to \mathcal{D}_{\mathit{fs}}\big(\{R, G, B, Y\}\big)$. We leave it at this stage to the reader to define an (inverse) map, in the opposite direction. Details will be provided in Section 3.

Stick breaking emerged in the description of stochastic processes, see [29] for an early source and [11] for an overview. The stick breaking isomorphism in (1) is applied to an iterated product (power) of spaces $(0,1)$ on the right-hand-side, without any dependencies. One can take for instance a (tensor) product of beta distributions on this product of $(0,1)$'s, and then transfer the result to a distribution on a space of the form $\mathcal{D}_{\mathit{fs}}(X)$ via stick breaking. In this way one obtains the (continuous) Dirichlet distribution (on discrete distributions) via multiple beta distributions and stick breaking. This is a known result — but not a very *well* known one — which we redescribe in Section 6 in the present setting.

Interestingly, the stick breaking construction can also be used for *infinite* products. It then yields an isomorphism $\mathcal{D}^{\infty}_{\mathit{fs}}(\mathbb{N}) \cong (0,1)^{\mathbb{N}}$, where we write $\mathcal{D}^{\infty}_{\mathit{fs}}$ for (discrete) distributions with infinite, full support. The above stick breaking construction is clearly repetitive, which suggests a coalgebraic structure. Indeed, as we shall see, the set of distributions $\mathcal{D}_{fs}^{\infty}(\mathbb{N})$ carries a coalgebra, which is even final.

It is especially this infinite form of stick breaking that is exploited in [29], and other sources like [6,11,24,25], to describe stochastic processes via infinite products followed by stick breaking. We give an impression of how this works, but only scratch the surface. The contribution of this paper lies in extracting the stick breaking operation from stochastic applications, in studying stick breaking on its own right, from a coalgebraic perspective, both in finite and infinite form, and then in re-applying the resulting insights in a few probabilistic illustrations.

The paper first fixes notation for discrete probability distributions, in order to introduce stick breaking in a coalgebraic setting, in Section 3. Then, after describing the essentials of multisets (bags) in Section 4, stick breaking is used to express multinomial draws from an urn in terms of successive binomial draws. This shows how drawing several balls from an urn with balls of multiple colours can be mimicked via urns with balls having only two colours (say black and white). This is *a priori* not entirely trivial.

The paper then moves on to continuous probability. It first shows how to express Dirichlet ditributions as parallel beta distributions, followed by stick breaking — in analogy with the connection between multinomials and binomials via stick breaking. Insiders of the field will probably say "sure, we are aware of such connections", but to (relative) outsiders they may provide useful insight. At this stage it is assumed that the reader has a basic level of familiarity with these standard distributions. Subsequently, the use of infinite stick breaking is illustrated for the definition of stochastic processes in terms of countably many parallel beta distributions. In this setting a mean is calculated. In the end it is shown that this mean arises by finality from a very simple coalgebraic construction.

2 Discrete probability distributions

Let X be an arbitrary set. There are two equivalent ways of describing (discrete, finite) probability distibutions on X.

- As finite convex combinations $r_1|x_1\rangle + \cdots r_n|x_n\rangle$ of elements $x_i \in X$, with probabilities $r_i \in [0,1]$ satisfying $\sum_i r_i = 1$.
- As functions $\omega\colon X \to [0,1]$ with finite support $\mathrm{supp}(\omega) \coloneqq \{x \in X \mid \omega(x) \neq 0\}$ and with $\sum_x \omega(x) = 1$.

We freely switch between these two descriptions. We write $\mathcal{D}(X)$ for the set of such distributions on X, and $\mathcal{D}_{fs}(X) \subseteq \mathcal{D}(X)$ for the subset of distributions with *full support*, that is, with $\mathrm{supp}(\omega) = X$. Thus, writing $\mathcal{D}_{fs}(X)$ only makes sense when the set X is finite.

This $\mathcal{D}$ is a monad on the category **Sets**. We make occasional use of this fact, so we do not spell out the details here; we refer to external sources instead, like [13,14].

We shall write $\mathcal{D}^\infty(X)$ for arbitrary functions $\omega\colon X \to [0,1]$ with $\sum_x \omega(x) = 1$. We then put no restriction on the support of ω, but it is not hard to show that when $\sum_x \omega(x) = 1$ the support is countable, or finite.

We write $\mathcal{D}^\infty_{\mathit{fs}}(X) \subseteq \mathcal{D}^\infty(X)$ for the subset of distributions with *full support*. This only makes sense if the set X is countable. We use it especially for $X = \mathbb{N}$.

3 Stick breaking

We fix a set A and consider the functor $A \times (-)\colon \mathbf{Sets} \to \mathbf{Sets}$. It is well known that the final coalgebra of this functor is the set $A^{\mathbb{N}}$ of infinite sequences of $(a_n)_{n\in\mathbb{N}}$ of elements $a_n \in A$.

Similarly, the functor $A + A \times (-)$ has the set $A^\infty := A^+ + A^{\mathbb{N}}$ of non-empty finite and infinite sequences as final coalgebra. Details can be found in any introductory text to coalgebra, see *e.g.* [1,13,17,27].

In the sequel we take $A = (0,1)$, the open unit interval $(0,1) \subseteq \mathbb{R}$ of numbers between zero and one. We introduce stick breaking first in the infinite case. Subsequently, the finite case is handled.

3.1 Infinite stick breaking

By finality we introduce a function $f\colon \mathcal{D}^\infty_{\mathit{fs}}(\mathbb{N}) \to (0,1)^{\mathbb{N}}$ in the following diagram.

$$\begin{array}{ccc}
(0,1) \times \mathcal{D}^\infty_{\mathit{fs}}(\mathbb{N}) & \xrightarrow{\;\mathrm{id}\times f\;} & (0,1) \times (0,1)^{\mathbb{N}} \\
{\scriptstyle \mathit{shift}}\uparrow & & \uparrow{\scriptstyle \cong\ \langle \mathit{head},\mathit{tail}\rangle} \\
\mathcal{D}^\infty_{\mathit{fs}}(\mathbb{N}) & \xrightarrow{\quad f \quad} & (0,1)^{\mathbb{N}}
\end{array} \tag{2}$$

The shift coalgebra on the left is defined as:

$$\mathit{shift}(\omega) := \left(\omega(0),\ \sum_{n\in\mathbb{N}} \frac{\omega(n+1)}{1-\omega(0)} \big| n \big\rangle \right). \tag{3}$$

This shift operation does three things: (1) it takes the head $\omega(0)$ of the infinite sequence $\omega = (\omega(0), \omega(1), \ldots)$; (2) it shifts the remaining tail one position forwards, so that $\omega(1)$ becomes the new head; (3) it renormalises this tail to a new distribution via division by $1 - \omega(0) = \sum_{n\geq 1} \omega(n)$.

Since each $\omega \in \mathcal{D}^\infty_{\mathit{fs}}(\mathbb{N})$ has full support, each probability $\omega(n)$ is non-zero, for $n \in \mathbb{N}$. But then none of these $\omega(n)$ can be equal to one. This ensures that the shift map is well-defined.

Proposition 1. *The function $f\colon \mathcal{D}^\infty_{\mathit{fs}}(\mathbb{N}) \to (0,1)^{\mathbb{N}}$ introduced in* (2) *by finality is an isomorphism. We shall write* $\mathrm{sb} = f^{-1}$ *for the inverse and call it (infinite)* stick breaking.

As a result, the shift coalgebra is also final — and thus an isomorphism.

Proof. Via commutation of Diagram (2) we get:

$$f(\omega) \;=\; \left(\omega(0),\, \frac{\omega(1)}{1-\omega(0)},\, \frac{\omega(2)}{1-\omega(0)-\omega(1)}, \ldots, \frac{\omega(i)}{1-\sum_{j<i}\omega(j)}, \ldots\right).$$

For instance, the second entry is obtained as:

$$\frac{\left(\frac{\omega(2)}{1-\omega(0)}\right)}{1-\frac{\omega(1)}{1-\omega(0)}} \;=\; \frac{\omega(2)}{1-\omega(0)-\omega(1)}.$$

In the other direction one obtains stick breaking as:

$$sb\Big(r_0, r_1, \ldots\Big) \coloneqq r_0\big|0\big\rangle + r_1(1-r_0)\big|1\big\rangle + \cdots + r_i \prod_{j<i}(1-r_j)\big|i\big\rangle + \cdots \tag{4}$$

If we abbreviate $\rho \coloneqq sb\big(r_0, r_1, \ldots\big)$ then we get as basic property, for each $i \in \mathbb{N}$

$$1 - \sum_{j \leq i} \rho(j) \;=\; \prod_{j \leq i}(1-r_j). \tag{5}$$

This follows by induction on i. The statement trivially holds for $i = 0$. Next,

$$\begin{aligned} 1 - \sum_{j \leq i+1} \rho(j) \;&=\; \left(1 - \sum_{j \leq i}\rho(j)\right) - \rho(i+1) \\ &\overset{\text{(IH)}}{=} \prod_{j \leq i}(1-r_j) - r_{i+1}\prod_{j \leq i}(1-r_j) \\ &=\; (1-r_{i+1})\prod_{j \leq i}(1-r_j) \;=\; \prod_{j \leq i+1}(1-r_j). \end{aligned}$$

We can now see that the sequence ρ forms a proper distribution:

$$\begin{aligned} \sum_{i \in \mathbb{N}} \rho(i) \;=\; \lim_{i \to \infty} \sum_{j \leq i}\rho(j) \;&=\; 1 - \lim_{i \to \infty}\, 1 - \sum_{j \leq i}\rho(j) \\ &\overset{(5)}{=} 1 - \lim_{i \to \infty} \prod_{j \leq i}(1-r_j) \;=\; 1 - 0 \;=\; 1. \end{aligned}$$

This works because an infinite product of numbers $s_i \in (0, 1)$ is zero.

It is not hard to see that these two functions $f\colon \mathcal{D}_{fs}^{\infty}(\mathbb{N}) \to (0,1)^{\mathbb{N}}$ and $sb\colon (0,1)^{\mathbb{N}} \to \mathcal{D}_{fs}^{\infty}(\mathbb{N})$ are each other's inverses. □

Example 2. Consider the infinite distribution:

$$\omega \;=\; \sum_{n \in \mathbb{N}} \tfrac{2}{5} \cdot \left(\tfrac{3}{5}\right)^n \big|n\big\rangle \;=\; \tfrac{2}{5}\big|0\big\rangle + \tfrac{6}{25}\big|1\big\rangle + \tfrac{18}{125}\big|2\big\rangle + \tfrac{54}{625}\big|3\big\rangle + \cdots$$

We can see that it is a distribution via the familiar formula:

$$\sum_{n\geq 0} r^n \;=\; \frac{1}{1-r} \qquad \text{for } r \in (0,1). \tag{6}$$

Then:

$$\sum_{n\geq 0} \omega(n) \;=\; \tfrac{2}{5}\cdot \sum_{n\geq 0} \left(\tfrac{3}{5}\right)^n \stackrel{(6)}{=} \tfrac{2}{5}\cdot \frac{1}{1-{}^3\!/_5} \;=\; \frac{2}{5-3} \;=\; 1.$$

The sequence of numbers in $(0,1)$ corresponding to ω is constant:

$$sb^{-1}(\omega) \;=\; (\tfrac{2}{5}, \tfrac{2}{5}, \tfrac{2}{5}, \ldots).$$

In general, for $r \in (0,1)$, we have $sb(r,r,r,\ldots) = \sum_{n\geq 0} r(1-r)^n|n\rangle$.

3.2 Finite stick breaking

Having seen the isomorphism $\mathcal{D}^{\infty}_{\mathit{fs}}(\mathbb{N}) \cong (0,1)^{\mathbb{N}}$ in Proposition 1 one wonders if it can be restricted to distributions with finite support. We write $\underline{n} = \{0,1,\ldots,n-1\}$, where $n \in \mathbb{N}$, for a chosen set with n elements.

We look at the set of distributions $\mathcal{D}_{\mathit{fs}}(\underline{n})$. Given a distribution $\omega \in \mathcal{D}_{\mathit{fs}}(\underline{n})$ we can apply a shift operation like in (3), to peel off the first element $\omega(0)$. However, what remains is a (full) distribution on $\underline{n-1}$. This gives a function $\mathcal{D}_{\mathit{fs}}(\underline{n}) \to (0,1)\times\mathcal{D}_{\mathit{fs}}(\underline{n-1})$, for $n > 0$. This is not a coalgebra, in the ordinary sense — but it may be understood as a coalgebra in dependent type theory.

We need a trick. We incorporate $\mathcal{D}_{\mathit{fs}}(\underline{n})$ into a subset of $\mathcal{D}^{\infty}(\mathbb{N})$, namely the subset were probabilities may be zero, but once they are zero, they remain zero in all subsequent positions. We use the following *ad hoc* notation.

$$\begin{aligned}\mathcal{D}^{\infty}_{\mathit{fs}<}(\mathbb{N}) \;&\coloneqq\; \Big\{\omega\colon \mathbb{N}\to[0,1) \;\Big|\; \textstyle\sum_n \omega(n) = 1 \text{ and } \forall n.\, \omega(n)=0 \Rightarrow \forall m>n.\, \omega(m)=0\Big\}\\ &\subseteq\; \mathcal{D}^{\infty}(\mathbb{N}).\end{aligned}$$

Notice that the 'shortest' list in $\mathcal{D}^{\infty}_{\mathit{fs}<}(\mathbb{N})$ is of the form $r|0\rangle + (1-r)|1\rangle$ for $r \in (0,1)$. For each $n > 1$ there is an inclusion $\mathcal{D}_{\mathit{fs}}(\underline{n}) \hookrightarrow \mathcal{D}^{\infty}_{\mathit{fs}<}(\mathbb{N})$.

We can now define a shift map of the following form, for $n > 0$.

$$\mathcal{D}^{\infty}_{\mathit{fs}<}(\mathbb{N}) \xrightarrow{\;\;\mathit{shift}\;\;} (0,1) + (0,1)\times\mathcal{D}^{\infty}_{\mathit{fs}<}(\mathbb{N})$$

This function is defined as:

$$\mathit{shift}(\omega) \;\coloneqq\; \begin{cases} r & \text{if } \omega = r|0\rangle + (1-r)|1\rangle \\ \left(\omega(0), \sum_n \frac{\omega(n+1)}{1-\omega(0)}|n\rangle\right) & \text{otherwise}\end{cases} \tag{7}$$

In the first case we have reached a distribution of minimal size. The second case is as in (3).

As mentioned in the beginning of this section, the set $(0,1)^{\infty} = (0,1)^{+} + (0,1)^{\mathbb{N}}$ of non-empty finite and infinite sequences of numbers in the open interval

$(0,1)$ forms a final coalgebra of the functor $(0,1) + (0,1) \times (-)$. By finality we thus get a map g in:

$$\begin{array}{ccc}
(0,1) + (0,1) \times \mathcal{D}^{\infty}_{\mathrm{fs}<}(\underline{n}) & \xrightarrow{\mathrm{id}+(\mathrm{id}\times g)} & (0,1) + (0,1) \times (0,1)^{\infty} \\
\Big\uparrow \scriptstyle{shift} & & \cong \Big\uparrow \scriptstyle{next} \\
\mathcal{D}^{\infty}_{\mathrm{fs}<}(\underline{n}) & \xrightarrow{\quad g \quad} & (0,1)^{\infty} = (0,1)^{+} + (0,1)^{\mathbb{N}}
\end{array} \tag{8}$$

It is not hard to see that the map g sends a distribution $\frac{1}{16}|0\rangle + \frac{1}{4}|1\rangle + \frac{3}{16}|2\rangle + \frac{1}{2}|3\rangle$ to the sequence $\langle \frac{1}{16}, \frac{4}{15}, \frac{3}{11} \rangle \in (0,1)^{\infty}$.

We now get the finite analogue of Proposition 1. The proof is essentially as in the infinite case, and is left to the reader.

Proposition 3. *For each $n > 1$ the function g defined in (8) restricts to a function $\mathcal{D}_{\mathrm{fs}}(\underline{n}) \to (0,1)^{+}$. In fact, it forms an isomorphism $g \colon \mathcal{D}_{\mathrm{fs}}(\underline{n}) \xrightarrow{\cong} (0,1)^{n-1}$. Its inverse $sb \colon (0,1)^{n-1} \xrightarrow{\cong} \mathcal{D}_{\mathrm{fs}}(\underline{n})$ will be called* stick breaking*. It is given by:*

$$\begin{array}{l}
sb\big(r_0, \ldots, r_{n-2}\big) \\
\quad = r_0\big|0\big\rangle + r_1(1-r_0)\big|1\big\rangle + r_2(1-r_1)(1-r_0)\big|2\big\rangle + \cdots + \\
\qquad\quad r_{n-2}(1-r_{n-3})\cdots(1-r_0)\big|n-2\big\rangle + (1-r_{n-2})\cdots(1-r_0)\big|n-1\big\rangle.
\end{array}$$

□

Example 4. For instance,

$$\begin{array}{rcl}
sb\big(\frac{1}{4}, \frac{1}{3}, \frac{3}{4}\big) & = & \frac{1}{4}|0\rangle + \frac{1}{4}|1\rangle + \frac{3}{8}|2\rangle + \frac{1}{8}|3\rangle \\
sb\big(\frac{7}{8}, \frac{2}{3}, \frac{3}{4}\big) & = & \frac{7}{8}|0\rangle + \frac{1}{12}|1\rangle + \frac{1}{32}|2\rangle + \frac{1}{96}|3\rangle.
\end{array}$$

Stickbreaking does *not* preserve convex combinations. For instance:

$$\begin{array}{rcl}
\frac{1}{4} \cdot sb\big(\frac{1}{4}, \frac{1}{3}, \frac{3}{4}\big) + \frac{3}{4} \cdot sb\big(\frac{7}{8}, \frac{2}{3}, \frac{3}{4}\big) & = & \frac{23}{32}|0\rangle + \frac{1}{8}|1\rangle + \frac{15}{128}|2\rangle + \frac{5}{128}|3\rangle \\
& \neq & \frac{23}{32}|0\rangle + \frac{21}{128}|1\rangle + \frac{45}{512}|2\rangle + \frac{15}{512}|3\rangle \\
& = & sb\big(\frac{1}{4} \cdot \langle \frac{1}{4}, \frac{1}{3}, \frac{3}{4} \rangle + \frac{3}{4} \cdot \langle \frac{7}{8}, \frac{2}{3}, \frac{3}{4} \rangle\big).
\end{array}$$

4 Multisets

A *multiset* (or *bag*) is a 'subset' except that elements may occur multiple times. We write a multiset on a set X also in ket form, as a finite formal sum $n_1|x_i\rangle + \cdots + n_k|x_k\rangle$ of elements $x_i \in X$ and natural numbers $n_i \in \mathbb{N}$. Such a multiset can equivalently be described as a function $\varphi \colon X \to \mathbb{N}$ with finite support. We write $\mathcal{M}(X)$ for the set of finite multisets on X, and $\mathcal{M}_{\mathrm{fs}}(X) \subseteq \mathcal{M}(X)$ for the subset of multisets with full support, that is, with $\varphi(x) \neq 0$ for each $x \in X$; again, this only makes sense when the set X is finite. The multiset operation $\mathcal{M}$ is a monad on **Sets**, like $\mathcal{D}$.

We associate several numbers with a multiset $\varphi \in \mathcal{M}(X)$.

- The *size* $\|\varphi\| \coloneqq \sum_{x\in X} \varphi(x)$ is the total number of elements in the multiset, including multiplicities.
- The *factorial* $\varphi\,!\!! \coloneqq \prod_{x\in X} \varphi(x)!$ is the product of (ordinary) factorials of the multiplicities.
- The *multinomial coefficient* is $(\!\varphi\!) \coloneqq \frac{\|\varphi\|!}{\varphi\,!\!!}$.

We write $\mathcal{M}[K](X) \subseteq \mathcal{M}(X)$ for the subset of multisets with size $K \in \mathbb{N}$.

4.1 Binomial and multinomial distributions

Drawing coloured balls from an urn is one of the most basic probabilistic models, see *e.g.* [18,21,23,26] and many other references. Here we look at draws with replacement, known as binomial draws (when there are two colours) or multinomial draws (when there are multiple colours).

For a fixed number $K \in \mathbb{N}$ we describe the familiar *binomial* distributions via a function:

$$[0,1] \xrightarrow{\;bn[K]\;} \mathcal{D}\big(\{0,1,\ldots,K\}\big)$$

It captures the probability of drawing $i \in \{0,1,\ldots,K\}$ black balls from an urn with (only) black and white balls, out of K independent draws, each with black ball probability $r \in [0,1]$. Thus:

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

There is a *multinomial* version which assigns a probability to a multiset of size K, as a draw (with replacement) of K-many balls from an urn with balls whose colours are described by a set X. The distribution of colours over the balls in the urn is captured abstractly via a distribution $\omega \in \mathcal{D}(X)$. Multinomial distributions will thus be described as a function:

$$\mathcal{D}(X) \xrightarrow{\;mn[K]\;} \mathcal{D}\big(\mathcal{M}[K](X)\big)$$

The definition is:

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

The binomial version is a special case, when X is a two-element set $\underline{2}$, via the isomorphisms $\mathcal{D}(\underline{2}) \cong [0,1]$ and $\mathcal{M}[K](\underline{2}) \cong \{0,1,\ldots,K\}$. For more information, see *e.g.* [15,16].

We present one result about multinomials. It is useful to recall as preparation for a similar but more complicated result later on.

Lemma 5. *Let $\omega \in \mathcal{D}(X)$ be an 'urn' and K the size of draws. The mean of the multinomial* $\mathrm{mn}[K](\omega)$ *is $K \cdot \omega$. Explicitly,*

$$\mathrm{mean}\Big(\mathrm{mn}[K](\omega)\Big) \coloneqq \sum_{\varphi\in\mathcal{M}[K](X)} \mathrm{mn}[K](\omega)(\varphi)\cdot\varphi \;=\; K\cdot\omega.$$

Strictly speaking, $K \cdot \omega$ is not a multiset, since we allow only natural numbers as multiplicities. But for a result like this one may wish to allow non-negative reals too. Once we do so, we can use inclusions $\mathcal{D}(Y) \hookrightarrow \mathcal{M}(Y)$ and see this result as an application of the multiplication map μ of the multiset monad $\mathcal{M}$, in:

$$\begin{array}{ccccc} \mathcal{D}(X) & \xrightarrow{mn[K]} & \mathcal{D}\big(\mathcal{M}[K](X)\big) & \hookrightarrow & \mathcal{M}\big(\mathcal{M}(X)\big) \\ & & & & \downarrow \mu \\ & & \xrightarrow{\quad K\cdot(-) \quad} & & \mathcal{M}(X) \end{array} \tag{9}$$

Proof. Fix an arbitrary element $y \in X$.

$$\begin{aligned} & \Big(mean\big(mn[K](\omega)\big)\Big)(y) \\ &= \sum_{\varphi \in \mathcal{M}[K](X)} mn[K](\omega)(\varphi) \cdot \varphi(y) \\ &= \sum_{\varphi \in \mathcal{M}[K](X),\, \varphi(y)\neq 0} \varphi(y) \cdot \frac{K!}{\prod_x \varphi(x)!} \cdot \prod\nolimits_x \omega(x)^{\varphi(x)} \\ &= \sum_{\varphi \in \mathcal{M}[K](X),\, \varphi(y)\neq 0} \frac{K \cdot (K-1)!}{(\varphi(y)-1)! \cdot \prod_{x\neq y} \varphi(x)!} \cdot \omega(y) \cdot \omega(y)^{\varphi(y)-1} \cdot \prod_{x \neq y} \omega(x)^{\varphi(x)} \\ &= K \cdot \omega(y) \cdot \sum_{\varphi \in \mathcal{M}[K-1](X)} \frac{(K-1)!}{\prod_x \varphi(x)!} \cdot \prod\nolimits_x \omega(x)^{\varphi(x)} \\ &= K \cdot \omega(y) \cdot \sum_{\varphi \in \mathcal{M}[K-1](X)} mn[K-1](\omega)(\varphi) \\ &= K \cdot \omega(y). \end{aligned}$$

□

5 Multinomials as iterated binomials

A simple question is: can we mimic a draw of multiple coloured balls from an urn in terms of draws of only two colours? More precisely, can we express a multinomial draw in terms of several binomial draws? We then encounter the problem that binomial draws use probabilities between zero and one and multinomials draws use distributions, as convex combinations. We show that stick breaking *sb* provides the connection. We first give a concrete formulation and then express it more abstractly.

Lemma 6. *Fix $n \geq 1$ and $K \geq 0$. For probabilities $\vec{r} = r_0, \ldots, r_{n-2} \in (0,1)^{n-1}$ and a multiset $\varphi = \sum_{i<n} k_i | i \rangle \in \mathcal{M}[K](\underline{n})$,*

$$\begin{aligned} mn[K]\big(sb(\vec{r})\big)(\varphi) \;=\; & bn[K](r_0)(k_0) \cdot bn[K-k_0](r_1)(k_1) \\ & \cdot \ldots \cdot bn[K - \textstyle\sum_{i<n-2} k_i](r_{n-2})(k_{n-2}). \end{aligned}$$

Notice that the last multiplicity $k_{n-1} = \varphi(n-1)$ is not used. It is superfluous if we know that the multiset has size K, since then $k_{n-1} = K - \sum_{i<n-1} k_i$.

Proof. One can use induction on n. When $n = 1$ the above equation formulates a binary multinomial as binomial, via the isomorphisms $\mathcal{D}(\underline{2}) \cong [0,1]$ and $\mathcal{M}[K](\underline{2}) \cong \{0,1,\ldots,K\}$. Concretely:

$$mn[K]\big(r|0\rangle + (1-r)|1\rangle\big)\big(k_0|0\rangle + k_1|1\rangle\big) \;=\; bn[K](r)(k_0).$$

Next, let $\varphi = \sum_{i \leq n} k_i|i\rangle \in \mathcal{M}[K](\underline{n+1})$ and $\vec{r} = r_0,\ldots,r_{n-1} \in (0,1)^n$ be given. We use a shifted multiset $\varphi' = \sum_{i<n-1} k_{i+1}|i\rangle$ of size $K-k_0$. Then:

$$\begin{array}{rl}
& bn[K](r_0)(k_0) \cdot bn[K-k_0](r_1)(k_1) \cdot \ldots \cdot bn[K-\sum_{i<n-1} k_i](r_{n-1})(k_{n-1}) \\
\overset{\text{(IH)}}{=} & bn[K](r_0)(k_0) \cdot mn[K-k_0]\big(sb(r_1,\ldots,r_{n-1})\big)(\varphi') \\
= & \displaystyle\binom{K}{k_0} \cdot r_0^{k_0} \cdot (1-r_0)^{K-k_0} \cdot \boldsymbol{(}\varphi'\boldsymbol{)} \cdot \prod_{i>0} sb(r_1,\ldots,r_{n-1})(i)^{k_i} \\
= & \displaystyle\frac{K!}{k_0! \cdot (K-k_0)!} \cdot \frac{(K-k_0)!}{k_1! \cdots k_{n-1}!} \cdot r_0^{k_0} \cdot \prod_{i>0} \Big(sb(r_1,\ldots,r_{n-1})(i) \cdot (1-r_0)\Big)^{k_i} \\
= & \boldsymbol{(}\varphi\boldsymbol{)} \cdot \prod_{i \geq 0} sb(r_0,\ldots,r_{n-1})(i)^{k_i} \\
= & mn[K]\big(sb(\vec{r})\big)(\varphi).
\end{array}$$

□

We reorganise this result a bit. For $K, n \in \mathbb{N}$ with $n > 0$ we define a set of sequences of natural numbers.

$$\mathcal{S}[K](n) \coloneqq \{(k_0,\ldots,k_{n-2}) \in \mathbb{N}^{n-1} \mid \forall i.\, k_i \leq K - \sum_{j<i} k_j\}.$$

Next we define the *sequential binomial* map $sbn[K] \colon (0,1)^{n-1} \to \mathcal{D}\big(\mathcal{S}[K](n)\big)$ by:

$$\begin{array}{rl}
sbn[K](\vec{r})(\vec{k}) \;=\; & bn[K](r_0)(k_0) \cdot bn[K-k_0](r_1)(k_1) \\
& \cdot \ldots \cdot bn[K-\sum_{i<n-2} k_i](r_{n-2})(k_{n-2}).
\end{array}$$

Theorem 7. *In the situation described above, multinomial distributions can be described as sequential binomial distributions via stick breaking, as in the following commuting diagram.*

$$\begin{array}{ccc}
\mathcal{D}\big(\mathcal{S}[K](n)\big) & \xrightarrow{\;\cong\;} & \mathcal{D}\big(\mathcal{M}[K](\underline{n})\big) \\
{\scriptstyle sbn[K]}\uparrow & & \uparrow{\scriptstyle mn[K]} \\
(0,1)^{n-1} & \underset{\cong}{\xrightarrow{\;sb\;}} & \mathcal{D}\big(\underline{n}\big)
\end{array}$$

Proof. This is just a fancy reformulation of Lemma 6. It uses the obvious isomorphism $\mathcal{S}[K](n) \xrightarrow{\cong} \mathcal{M}[K](\underline{n})$, given by $(k_0,\ldots,k_{n-2}) \mapsto \sum_{i<n-1} k_i|i\rangle + (K-\sum_i k_i)|n-1\rangle$, at the top, together with the functoriality of $\mathcal{D}$. □

6 Dirichlet via parallel Beta's

This section describes an application of stick breaking in continuous probability theory. It reformulates the famous Dirichlet distribution in terms of parallel beta distributions, with stick breaking forming the connection. This is similar to the result in the previous section, since beta distributions can be understood as binary versions of Dirichlet distributions — just like binomials being binary versions of multinomials. For background information on the beta and Dirichlet distributions we refer to standard textbooks, like [2,3,8,19,30].

We shall describe these continuous distributions via the Giry monad $\mathcal{G}$, which generalises the discrete probability monad $\mathcal{D}$, see [12,14,22] for details. We shall use continuous probability distributions on subsets $S \subseteq \mathbb{R}^n$, given by a probability density function (pdf) $f\colon S \to \mathbb{R}_{\geq 0}$, satisfying $\int f(x)\,\mathrm{d}x = 1$. The distribution itself is given by a mapping from the Borel σ-algebra Σ_S of measurable subsets of S, to $[0,1]$. Thus, it is the mapping on measurable subsets $M \subseteq S$,

$$M \longmapsto \int_{x \in M} f(x)\,\mathrm{d}x.$$

We write $\mathcal{G}(S)$ for the set of such distributions. For $\phi \in \mathcal{G}(S)$ and $\chi \in \mathcal{G}(T)$ there is a parallel product $\phi \otimes \chi \in \mathcal{G}(S \times T)$ determined by $(\phi \otimes \chi)(M \times N) = \phi(M) \cdot \chi(N)$, for measurable subsets $M \subseteq S$, $N \subseteq T$.

We illustrate this for the beta distributions on $(0,1)$, which we describe as parameterised by numbers $a, b \in \mathbb{N}_{>0}$. This can be generalised to more general numbers, but we don't need that here. The pdf $\mathit{pbf}_{\mathit{Beta}}(a,b)\colon (0,1) \to \mathbb{R}_{\geq 0}$ is given by:

$$\mathit{pbf}_{\mathit{Beta}}(a,b)(r) \coloneqq \frac{r^{a-1} \cdot (1-r)^{b-1}}{B(a,b)} \quad \text{where} \quad B(a,b) = \frac{(a-1)! \cdot (b-1)!}{(a+b-1)!}. \tag{10}$$

The Dirichlet distribution takes the form of a map:

$$\mathcal{M}_{\mathit{fs}}(\underline{n}) \xrightarrow{\quad \mathit{Dir} \quad} \mathcal{G}\big(\mathcal{D}_{\mathit{fs}}(\underline{n})\big). \tag{11}$$

For a multiset $\psi \in \mathcal{M}_{\mathit{fs}}(\underline{n})$ we describes its pdf $\mathcal{D}_{\mathit{fs}}(\underline{n}) \to \mathbb{R}_{\geq 0}$ as:

$$\mathit{pbf}_{\mathit{Dir}(\psi)}(\omega) \coloneqq \frac{(\|\psi\|-1)!}{(\psi-\mathbf{1})!\!\!\!\circ} \cdot \prod_{i \in \underline{n}} \omega(i)^{\psi(i)-1} \quad \text{where} \quad \mathbf{1} = \sum_{i \in \underline{n}} 1|\,i\,\rangle.$$

This looks very much like the multinomial distribution $\mathit{mn}[K]$. Indeed, there is a close connection: if we view the multinomial as a map $\mathit{mn}[K]\colon \mathcal{D}_{\mathit{fs}}(\underline{n}) \to \mathcal{D}(\mathcal{M}[K](\underline{n})) \cong \mathcal{G}(\mathcal{M}[K](\underline{n}))$ then Dirichlet is its *dagger* [5,4,10] in the opposite direction (11), using a uniform prior. Details will be elaborated elsewhere. A further basic fact is that the Kleisli composition 'multinomial after Dirichlet' yields Pólya distributions [21].

Our focus lies on the theorem below that expresses the Dirichlet distribution as parallel product $\otimes$ of beta's, connected via stick breaking. This is a known 'folklore' result, for which it is hard to find a precise reference and/or formulation, but see [11, §3.1] for a brief description. As an aside, there is also a way to express Dirichlet via gamma distributions that is more familiar, see *e.g.* [30, 7.7.1] or [7, Prop. 4.1].

Here we can precisely formulate Dirichlet via beta's because we have explicitly identified the (finite) stick breaking isomorphism $sb\colon (0,1)^{n-1} \xrightarrow{\cong} \mathcal{D}_{fs}(\underline{n})$. The formulation below uses functoriality of Giry $\mathcal{G}$.

Theorem 8. *For $n > 0$ and $\psi \in \mathcal{M}(\underline{n})$,*

$$\begin{aligned} Dir(\psi) \;=\; \mathcal{G}(sb)\Big(& Beta\big(\psi(0), \textstyle\sum_{i>0}\psi(i)\big) \otimes Beta\big(\psi(1), \textstyle\sum_{i>1}\psi(i)\big) \otimes \cdots \\ & \cdots \otimes Beta\big(\psi(n-3), \psi(n-2)+\psi(n-1)\big) \otimes Beta\big(\psi(n-2), \psi(n-1)\big)\Big). \end{aligned}$$

Proof. One proceeds like in the proof of Lemma 6, in combination with integration by substitution. We give an exemplaric proof, for $n = 3$, illustrating how this works.

The Dirichlet distribution involves, in this case, an integral over $\mathcal{D}_{fs}(\underline{3})$. This means that we integrate over $(0,1)$, say with a variable s_0, and then over $(0, 1-s_0)$, say with s_1, and then use $s_2 = 1 - s_0 - s_1$. We thus restrict the inverse of the stick breaking isomorphism $sb\colon (0,1)^2 \xrightarrow{\cong} \mathcal{D}_{fs}(\underline{3})$ to an isomorphism:

$$D_2 \coloneqq \{(s_0, s_1) \mid s_0 \in (0,1), s_1 \in (0, 1-s_0)\} \xrightarrow[\cong]{h} (0,1)^2$$

There is an isomorphism $\mathcal{D}_{fs}(\underline{3}) \cong D_2$ via dropping the last number. This function $h = (h_0, h_1)$ is thus given by:

$$h(s_0, s_1) \;=\; (s_0, \tfrac{s_1}{1-s_0}).$$

In order to do (multidimensional) integration by substitution we need the determinant of the matrix of partial derivatives of h. This is:

$$\begin{vmatrix} \dfrac{\partial h_0}{\partial s_0}(\vec{s}) & \dfrac{\partial h_0}{\partial s_1}(\vec{s}) \\ \dfrac{\partial h_1}{\partial s_0}(\vec{s}) & \dfrac{\partial h_1}{\partial s_1}(\vec{s}) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \dfrac{s_1}{1-s_0} & \dfrac{1}{1-s_0} \end{vmatrix} = \frac{1}{1-s_0}. \qquad (*)$$

We are now ready to prove the equation in the theorem, for $n = 3$. We fix $\psi \in \mathcal{M}_{fs}(\underline{3})$. Let $M \subseteq \mathcal{D}_{fs}(\underline{3})$ be an arbitrary measurable subset; we identify it with $\overline{M} \subseteq D_2$ when needed, via the isomorphism $\mathcal{D}_{fs}(\underline{3}) \cong D_2$ described above.

$$
\begin{aligned}
&\mathcal{G}(sb)\Big(Beta\big(\psi(0), \psi(1){+}\psi(2)\big) \otimes Beta\big(\psi(1), \psi(2)\big)\Big)(M) \\
&= \int_{(r_0,r_1)\in sb^{-1}(M)} pbf_{Beta}\big(\psi(0), \psi(1){+}\psi(2)\big)(r_0) \cdot pbf_{Beta}\big(\psi(1), \psi(2)\big)(r_1)\, \mathrm{d}r_0, r_1 \\
&\overset{(10)}{=} \int_{(r_0,r_1)\in h(\overline{M})} \frac{r_0^{\psi(0)-1} \cdot (1{-}r_0)^{\psi(1)+\psi(2)-1}}{B(\psi(0), \psi(1){+}\psi(2))} \cdot \frac{r_1^{\psi(1)-1} \cdot (1{-}r_1)^{\psi(2)-1}}{B(\psi(1), \psi(2))}\, \mathrm{d}r_0, r_1 \\
&= \int_{(s_0,s_1)\in\overline{M}} \frac{s_0^{\psi(0)-1} \cdot (1{-}s_0)^{\psi(1)+\psi(2)-1}}{B(\psi(0), \psi(1){+}\psi(2))} \\
&\qquad \cdot \frac{\big(\frac{s_1}{1-s_0}\big)^{\psi(1)-1} \cdot \big(1{-}\frac{s_1}{1-s_0}\big)^{\psi(2)-1}}{B(\psi(1), \psi(2))} \cdot \frac{1}{1{-}s_0}\, \mathrm{d}s_0, s_1 \quad \text{via substitution, using } (*) \\
&\overset{(10)}{=} \int_{(s_0,s_1)\in\overline{M}} \frac{s_0^{\psi(0)-1}}{\frac{(\psi(0)-1)!\cdot(\psi(1)+\psi(2)-1)!}{(\psi(0)+\psi(1)+\psi(2)-1)!}} \cdot \frac{s_1^{\psi(1)-1} \cdot \big(1{-}s_0{-}s_1\big)^{\psi(2)-1}}{\frac{(\psi(1)-1)!\cdot(\psi(2)-1)!}{(\psi(1)+\psi(2)-1)!}}\, \mathrm{d}s_0, s_1 \\
&= \int_{\omega\in M} \frac{(\|\psi\|{-}1)!}{(\psi(0)-1)! \cdot (\psi(1)-1)! \cdot (\psi(2)-1)!} \cdot \prod_{i\in\underline{3}} \omega(i)^{\psi(i)-1}\, \mathrm{d}\omega \\
&= \int_{\omega\in M} \frac{(\|\psi\|{-}1)!}{(\psi{-}\mathbf{1})!!} \cdot \prod_{i\in\underline{3}} \omega(i)^{\psi(i)-1}\, \mathrm{d}\omega \\
&= Dir(\psi)(M).
\end{aligned}
$$

The proof in general, for arbitrary $n > 0$, works in the same way, but involves much more book keeping. □

7 Infinite stick breaking and beta distributions

In the literature on stochastic processes infinite stick breaking $sb\colon (0,1)^{\mathbb{N}} \xrightarrow{\cong} \mathcal{D}^{\infty}_{fs}(\mathbb{N})$ from Proposition 1 is used as construction to produce (continuous) distributions on (discrete, infinite) distributions in $\mathcal{D}^{\infty}_{fs}(\mathbb{N})$. For numbers $a_n, b_n \in \mathbb{N}_{>0}$ one can define:

$$
sbB(a,b) \coloneqq \mathcal{G}(sb)\left(\bigotimes_{n\in\mathbb{N}} Beta(a_n, b_n)\right) \in \mathcal{G}\big(\mathcal{D}^{\infty}_{fs}(\mathbb{N})\big). \tag{12}
$$

The abbreviation *sbB* stands for 'stick break Beta'; it is described as 'stick-breaking prior' in [11, §1.1]. When we pull out the parameters we get a function:

$$
\big(\mathbb{N}_{>0}\big)^{\mathbb{N}} \times \big(\mathbb{N}_{>0}\big)^{\mathbb{N}} \xrightarrow{\quad sbB \quad} \mathcal{G}\big(\mathcal{D}^{\infty}_{fs}(\mathbb{N})\big) \tag{13}
$$

Examples of such stochastic processes $sbB(a,b)$ are Dirichlet-Poisson [6,9,20] and Pitman-Yor [24,25]. For instance, in the Dirichlet-Poisson case the sequence

a is contantly one, and the sequence b is also constant, determined by a parameter. For Pitman-Yor only a is constant. These stochastic processes are used for infinite mixture models, as "stick breaking priors", see [11] for an overview.

As an aside, the probabilities in distributions in $\mathcal{D}_{fs}^{\infty}(\mathbb{N})$ are sometimes used in descending order, see *e.g.* [20, Appendix], so that what is commonly called Dirichlet-Poisson is a quotient of our general formulation (12). However, here we abstract away from such matters and will simply work with the above formulation.

We concentrate on one small thing, namely computing the mean of a stick break beta process (12). This allows us to conclude this article with a coalgebraic observation. Thus, in the style of Diagram (9) our goal is to describe the composite:

$$\begin{array}{ccccc} \big(\mathbb{N}_{>0}\big)^{\mathbb{N}} \times \big(\mathbb{N}_{>0}\big)^{\mathbb{N}} & \xrightarrow{sbB} & \mathcal{G}\big(\mathcal{D}_{fs}^{\infty}(\mathbb{N})\big) & \hookrightarrow & \mathcal{G}\big(\mathcal{G}(\mathbb{N})\big) \\ & & & & \downarrow{\scriptstyle \mu} \\ & & & & \mathcal{G}(\mathbb{N}) \end{array} \tag{14}$$

where μ is the multiplication of the Giry monad. Interestingly, the outcome is a discrete distribution on $\mathbb{N}$.

We first observe that the mean can also be computed as Kleisli extension, which we write as $\gg\!=$. Indeed:

$$\begin{aligned} mean\big(sbB(a,b)\big) &= \mu\left(\mathcal{G}(sb)\left(\bigotimes_{n\in\mathbb{N}} Beta(a_n,b_n)\right)\right) \\ &= sb \gg\!= \left(\bigotimes_{n\in\mathbb{N}} Beta(a_n,b_n)\right). \end{aligned}$$

We first calculate the latter expression in the finite case. For instance, at position $0 \in \underline{3}$ of the distribution in $\mathcal{D}_{fs}(\underline{3})$ one has:

$$\begin{aligned} &\Big(sb \gg\!= \Big(Beta(a_0,b_0) \otimes Beta(a_1,b_1)\Big)\Big)(0) \\ &= \int_0^1 \int_0^1 sb(r_0,r_1)(0) \cdot pbf_{Beta}(a_0,b_0)(r_0) \cdot pbf_{Beta}(a_1,b_1)(r_1)\,\mathrm{d}r_1\,\mathrm{d}r_0 \\ &= \int_0^1 r_0 \cdot \frac{r_0^{a_0-1}\cdot(1-r_0)^{b_0-1}}{B(a_0,b_0)} \cdot \left(\int_0^1 pbf_{Beta}(a_1,b_1)(r_1)\,\mathrm{d}r_1\right)\mathrm{d}r_0 \\ &= \frac{B(a_0+1,b_0)}{B(a_0,b_0)} \overset{(10)}{=} \frac{a_0!\cdot(b_0-1)!}{(a_0+b_0)!}\cdot\frac{(a_0+b_0-1)!}{(a_0-1)!\cdot(b_0-1)!} = \frac{a_0}{a_0+b_0}. \end{aligned}$$

Similarly, at position 1,

$$\begin{aligned}
&\Big(sb \gg\!= \Big(\mathit{Beta}(a_0, b_0) \otimes \mathit{Beta}(a_1, b_1)\Big)\Big)(1) \\
&= \int_0^1 \int_0^1 sb(r_0, r_1)(1) \cdot \mathit{pbf}_{\mathit{Beta}}(a_0, b_0)(r_0) \cdot \mathit{pbf}_{\mathit{Beta}}(a_1, b_1)(r_1)\, \mathrm{d}r_1\, \mathrm{d}r_0 \\
&= \int_0^1 (1-r_0) \cdot \frac{r_0^{a_0-1} \cdot (1-r_0)^{b_0-1}}{B(a_0, b_0)} \cdot \int_0^1 r_1 \cdot \frac{r_1^{a_1-1} \cdot (1-r_1)^{b_1-1}}{B(a_1, b_1)}\, \mathrm{d}r_1\, \mathrm{d}r_0 \\
&= \frac{B(a_0, b_0+1)}{B(a_0, b_0)} \cdot \frac{B(a_1+1, b_1)}{B(a_1, b_1)} \;=\; \frac{b_0}{a_0+b_0} \cdot \frac{a_1}{a_1+b_1}.
\end{aligned}$$

Thus, Kleisli extension $\gg\!=$ gives the following distribution on $\underline{3}$.

$$\begin{aligned}
&sb \gg\!= \Big(\mathit{Beta}(a_0, b_0) \otimes \mathit{Beta}(a_1, b_1)\Big) \\
&= \frac{a_0}{a_0+b_0}\big|0\big\rangle + \frac{a_1 b_0}{(a_0+b_0)(a_1+b_1)}\big|1\big\rangle + \frac{b_0 b_1}{(a_0+b_0)(a_1+b_1)}\big|2\big\rangle.
\end{aligned}$$

This reveals the pattern. It can be extended to infinity.

Lemma 9. *For sequences $a, b \in \big(\mathbb{N}_{>0}\big)^{\mathbb{N}}$ the mean of stick-break-beta yields the following distribution in $\mathcal{D}^{\infty}_{\mathit{fs}}(\mathbb{N})$.*

$$\mathit{mean}\big(\mathit{sbB}(a, b)\big) \;=\; sb \gg\!= \left(\bigotimes_{n\in\mathbb{N}} \mathit{Beta}(a_n, b_n)\right) \;=\; \sum_{n\in\mathbb{N}} \frac{a_n \prod_{i<n} b_i}{\prod_{i\leq n}(a_i+b_i)}\,\big|n\big\rangle. \quad \square$$

For instance, for Poisson-Dirichlet we have $a_n = 1$ and $b_n = t$, where $t \in \mathbb{N}_{>0}$ is a parameter. The resulting mean is the infinite discrete distribution:

$$\sum_{n\in\mathbb{N}} \frac{t^{n-1}}{(1+t)^n}\,\big|n\big\rangle \;=\; \frac{1}{t}\sum_{n\in\mathbb{N}} \left(\frac{t}{1+t}\right)^n \big|n\big\rangle$$

We conclude by returning to a coalgebraic narrative. It turns out that the non-entirely trivial distribution in Lemma 9 can be obtained by finality from a completely trivial and standard coalgebra, involving the derivative a' of a sequence/stream a, see [28] for many more examples.

Proposition 10. *Consider the finality diagram:*

$$\begin{array}{ccc}
(0,1) \times \Big(\big(\mathbb{N}_{>0}\big)^{\mathbb{N}} \times \big(\mathbb{N}_{>0}\big)^{\mathbb{N}}\Big) & \xrightarrow{\;\mathrm{id}\times h\;} & (0,1) \times \mathcal{D}^{\infty}_{\mathit{fs}}(\mathbb{N}) \\
\uparrow c & & \cong \uparrow \mathit{shift} \\
\big(\mathbb{N}_{>0}\big)^{\mathbb{N}} \times \big(\mathbb{N}_{>0}\big)^{\mathbb{N}} & \xrightarrow{\;\;h\;\;} & \mathcal{D}^{\infty}_{\mathit{fs}}(\mathbb{N})
\end{array}$$

The coalgebra c on the left is defined as:

$$c(a, b) \coloneqq \big(\tfrac{a_0}{a_0+b_0},\, a',\, b'\big) \qquad \textit{where} \qquad \begin{cases} a'_n = a_{n+1} \\ b'_n = b_{n+1}. \end{cases}$$

The function $h\colon \big(\mathbb{N}_{>0}\big)^{\mathbb{N}} \times \big(\mathbb{N}_{>0}\big)^{\mathbb{N}} \to \mathcal{D}_{fs}^{\infty}(\mathbb{N})$ *obtained by finality is then the mean of stick-break-Beta, as described in Lemma 9.*

Proof. We recall that the *shift* coalgebra (3), in the rectangle on the right, is final, by Proposition 1. Let's write:

$$h(a,b) \;=\; mean\big(sbB(a,b)\big) \;=\; \sum_{n\in\mathbb{N}} \frac{a_n \prod_{i<n} b_i}{\prod_{i\leq n}(a_i + b_i)}\,\big|\,n\,\big\rangle.$$

It suffices to show that this h makes the above rectangle commute. We look at the first and second projections separately.

$$\big(\pi_1 \circ shift \circ h\big)(a,b) \overset{(3)}{=} h(a,b)(0) \;=\; \frac{a_0}{a_0 + b_0} \;=\; \big(\pi_1 \circ c\big)(a,b).$$

And:

$$\begin{aligned}
\big(\pi_2 \circ shift \circ h\big)(a,b) \overset{(3)}{=} \sum_{n\in\mathbb{N}} \frac{h(a,b)(n+1)}{1 - h(a,b)(0)}\big|\,n\,\big\rangle &= \sum_{n\in\mathbb{N}} \frac{\frac{a_{n+1}\prod_{i<n+1} b_i}{\prod_{i\leq n+1}(a_i+b_i)}}{1 - \frac{a_0}{a_0+b_0}}\big|\,n\,\big\rangle \\
&= \sum_{n\in\mathbb{N}} \frac{\frac{a_{n+1}\prod_{i<n+1} b_i}{\prod_{i\leq n+1}(a_i+b_i)}}{\frac{b_0}{a_0+b_0}}\big|\,n\,\big\rangle \\
&= \sum_{n\in\mathbb{N}} \frac{a_{n+1}\prod_{0<i<n+1} b_i}{\prod_{0<i\leq n+1}(a_i+b_i)}\big|\,n\,\big\rangle \\
&= \sum_{n\in\mathbb{N}} \frac{a'_n\prod_{i<n} b'_i}{\prod_{i\leq n}(a'_i+b'_i)}\big|\,n\,\big\rangle \\
&= h(a', b') \\
&= \big(h \circ \pi_2 \circ c\big)(a,b).
\end{aligned}$$

□

8 Concluding remarks

This paper extracts stick breaking from stochastic process theory and investigates it in a coalgebraic setting. This works smoothly for infinite stick breaking, yielding a new description $\mathcal{D}_{fs}^{\infty}(\mathbb{N})$ of the final coalgebra of the functor $(0,1)\times(-)$. In the finite case, the coalgebraic treatment of stick breaking is a bit artificial. Nevertheless, the following two stick breaking isomorphisms are both fundamental and useful.

$$(0,1)^{n-1} \xrightarrow{\;\cong\;} \mathcal{D}_{fs}(\underline{n}) \qquad\text{and}\qquad (0,1)^{\mathbb{N}} \xrightarrow{\;\cong\;} \mathcal{D}_{fs}^{\infty}(\mathbb{N}).$$

This usefulness has been illustrated by relating multinomials to iterated binomials and by relating Dirichlet to parallel Beta's. Also, one, coalgebraic, aspect of the use of infinite stick breaking in stochastic processes has been elaborated, namely the computation of the mean, via finality. This area of stochastic processes may benefit also in other ways from coalgebraic techniques.

Acknowledgements

Thanks are due to the anonymous reviewers for their useful feedback.

References

1. J. Adámek. Introduction to coalgebra. *Theory and Appl. of Categories*, 14(8):157–199, 2005.
2. J. Bernardo and A. Smith. *Bayesian Theory.* John Wiley & Sons, 2000. Available via `https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316870`. `doi:10.1002/9780470316870`.
3. P. Billingsley. *Probability and Measure.* Wiley-Interscience, New York, 1995.
4. K. Cho and B. Jacobs. Disintegration and Bayesian inversion via string diagrams. *Math. Struct. in Comp. Sci.*, 29(7):938–971, 2019. `doi:10.1017/s0960129518000488`.
5. F. Clerc, F. Dahlqvist, V. Danos, and I. Garnier. Pointless learning. In J. Esparza and A. Murawski, editors, *Foundations of Software Science and Computation Structures*, number 10203 in Lect. Notes Comp. Sci., pages 355–369. Springer, Berlin, 2017. `doi:10.1007/978-3-662-54458-7_21`.
6. H. Crane. The ubiquitous Ewens sampling formula. *Statistical Science*, 31(1):1–19, 2016. `doi:10.1214/15-STS529`.
7. V. Danos and I. Garnier. Dirichlet is natural. In D. Ghica, editor, *Math. Found. of Programming Semantics*, number 319 in Elect. Notes in Theor. Comp. Sci., pages 137–164. Elsevier, Amsterdam, 2015.
8. S. Feng. *The Poisson-Dirichlet Distribution and Related Topics.* Probability and its Applications. Springer-Verlag, Berlin, 2010.
9. T. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973. `doi:10.1214/aos/1176342360`.
10. T. Fritz. A synthetic approach to Markov kernels, conditional independence, and theorems on sufficient statistics. *Advances in Math.*, 370:107239, 2020. `doi:10.1016/J.AIM.2020.107239`.
11. H. Ishwaran and L. James. Gibbs sampling methods for stick-breaking priors. *Journ. of the Amer. Statistical Assoc.*, 96(453):161–173, 2001. `doi:10.1198/016214501750332758`.
12. B. Jacobs. Measurable spaces and their effect logic. In *Logic in Computer Science.* IEEE, Computer Science Press, 2013.
13. B. Jacobs. *Introduction to Coalgebra. Towards Mathematics of States and Observations.* Number 59 in Tracts in Theor. Comp. Sci. Cambridge Univ. Press, 2016.
14. B. Jacobs. From probability monads to commutative effectuses. *Journ. of Logical and Algebraic Methods in Programming*, 94:200–237, 2018. `doi:10.1016/j.jlamp.2016.11.006`.
15. B. Jacobs. From multisets over distributions to distributions over multisets. In *Logic in Computer Science.* IEEE, Computer Science Press, 2021. `doi:10.1109/lics52264.2021.9470678`.
16. B. Jacobs. Urns & tubes. *Compositionality*, 2022, to appear. See `https://arxiv.org/abs/2110.02155`.
17. B. Jacobs and J. Rutten. A tutorial on (co)algebras and (co)induction. In D. Sangiorgi and J. Rutten, editors, *Advanced topics in bisimulation and coinduction*, number 52 in Tracts in Theor. Comp. Sci., pages 38–99. Cambridge Univ. Press, 2011.

18. N. Johnson and S. Kotz. *Urn models and their application: An approach to modern discrete probability theory*. John Wiley, 1977.
19. O. Kallenberg. *Foundations of Modern Probability*. Number 99 in Probability Theory and Stochastic Modelling. Springer, 2021.
20. J. Kingman. Random partitions in population genetics. *Proc. Royal Soc.*, Series A, 361:1–20, 1978. doi:10.1098/rspa.1978.0089.
21. H. Mahmoud. *Pólya Urn Models*. Chapman and Hall, 2008.
22. P. Panangaden. *Labelled Markov Processes*. Imperial College Press, London, 2009.
23. H. Pishro-Nik. *Introduction to probability, statistics, and random processes*. Kappa Research LLC, 2014. Available at https://www.probabilitycourse.com.
24. J. Pitman. Random discrete distributions invariant under size-biased permutation. *Adv. in Applied Probability*, 28(2):525–539, 1995. doi:10.2307/1428070.
25. J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2):855–900, 1997. doi:10.1214/aop/1024404422.
26. S. Ross. *A first course in probability*. Pearson Education, 10^{th} edition, 2018.
27. J. Rutten. Universal coalgebra: a theory of systems. *Theor. Comp. Sci.*, 249:3–80, 2000. doi:10.1016/S0304-3975(00)00056-6.
28. J. Rutten. A coinductive calculus of streams. *Math. Struct. in Comp. Sci.*, 15(1):93–147, 2005. doi:10.1017/S0960129504004517.
29. J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994. doi:10.21236/ada238689.
30. S. Wilks. *Mathematical Statistics*. Wiley, 1962.