Stick Breaking, in Coalgebra & Probability

Radboud University Nijmegen
CMCS, April 3, 2022

Bart Jacobs
bart@cs.ru.nl

Outline

Where we are, so far

Background

- Stochastic processes are probability distribution on infinite products like $X^{\mathbb{N}}$, for some set X.
  - used e.g. for infinitely many coin throws — for $X = 2 = \{0, 1\}$
  - or as priors in clustering, with an unbounded number of clusters
- For some of these stochastic processes the definition involves a repetitive construction called stick breaking
  - first in (Sethuraman 1994)
  - e.g. (Ishwaran & James 2001) speaks of stick breaking priors
  - this raised my "coalgebraic antenna"
- The CMCS paper isolates stick breaking as a separate construction and puts it in proper coalgebraic perspective.

Probability distributions (discrete)

- Example distribution on four colours, notation with ket's:

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

- We write $\mathcal{D}_{fs}$ for set of distributions (on these colours), with full support

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

- Alternatively, as simplex of dimenstion three:

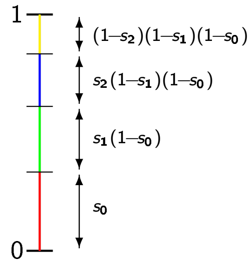
$$\mathcal{D}_{fs}(\{R,G,B,Y\}) \xrightarrow{\cong} \Big\{ (r_0,r_1,r_2) \in (0,1)^3 \;\Big|\; r_0+r_1+r_2 < 1 \Big\} \overset{\text{proper subset}}{\hookrightarrow} (0,1)^3$$

Stick breaking, finite example

Fact. There is a stick breaking isomorphism $\mathcal{D}_{fs}(\{R,G,B,Y\} \cong (0,1)^3$

Construction, starting from $(s_0,s_1,s_2) \in (0,1)^3$

- Take a stick of lenght 1, as on the right:
- Paint the lower part/proportion s_0 red, leaving an unpainted part of length $1-s_0$
- Paint the s_1 proportion of it green. The painted part then is $s_1(1-s_0)$ long. The unpainted part is $(1-s_2)(1-s_0)$
- Paint the s_2-proportion of the remainder blue
- The final remainder is then of length $(1-s_2)(1-s_2)(1-s_0)$; it becomes yellow



The result is a colour distribution:

$$s_0\big|\,R\,\big\rangle + s_1(1-s_0)\big|\,G\,\big\rangle + s_2(1-s_1)(1-s_0)\big|\,B\,\big\rangle + (1-s_2)(1-s_1)(1-s_0)\big|\,Y\,\big\rangle$$

Overview: two stick breaking isomorphisms

- Stick breaking sb can be done finitely and infinitely many times
- This gives isomorphisms:

$$(0,1)^{n-1} \xrightarrow[\cong]{sb} \mathcal{D}_{fs}(\{0,1,\ldots,n-1\})$$

$$(0,1)^{\mathbb{N}} \xrightarrow[\cong]{sb} \mathcal{D}_{fs}^{\infty}(\mathbb{N})$$

where:

- $\mathcal{D}_{fs}$ is used for distributions with finite full support
- $\mathcal{D}_{fs}^{\infty}$ is for distributions with infinite full support

Where we are, so far

Final coalgebras

- Fix a set A and recall the basic facts:
  - $A^{\mathbb{N}}$ is/carries final coalgebra for the functor $A \times (-)$ on Sets
  - $A^{\infty} := A^{+} + A^{\mathbb{N}}$ is final for $A + A \times (-)$ — with non-empty lists
- Our first aim is to prove $\mathcal{D}^{\infty}_{fs}(\mathbb{N}) \cong (0,1)^{\mathbb{N}}$
  - strategy: show that $\mathcal{D}^{\infty}_{fs}(\mathbb{N})$ carries final coalgebra for $(0,1) \times (-)$

Shift coalgebra on distributions

Via finality we can define $f\colon \mathcal{D}^{\infty}_{fs}(\mathbb{N}) \to (0,1)^{\mathbb{N}}$ in:

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

where:

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

Three things happen, for $\omega \in \mathcal{D}^{\infty}_{fs}(\mathbb{N})$,

- head $\omega(0)$ is taken off
- tail is shifted forward — like in a coalgebraic derivative
- this tail is renormalised to a distribution — using fulness

Infinite stick breaking

Theorem

The map $f\colon \mathcal{D}^{\infty}_{fs}(\mathbb{N}) \to (0,1)^{\mathbb{N}}$ by finality is an isomorphism, with stick breaking as inverse:

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

Example

The infinite distribution:

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

arises as $\omega = sb\left(\tfrac{2}{5}, \tfrac{2}{5}, \tfrac{2}{5}, \ldots\right)$.

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

Finite stick breaking is more subtle

Write $\underline{n} := \{0, 1, \ldots, n-1\}$ for a generic n-element set

The fullness requirement breaks coalgebra shape for shift:

$$\mathcal{D}(\underline{n}) \xrightarrow{\;shift\;} (0,1) \times \mathcal{D}(\underline{n-1})$$

Trick: use "once zero, forever zero" in tail, in ad hoc notation:

$$\mathcal{D}^{\infty}_{fs<}(\mathbb{N}) := \{\omega\colon \mathbb{N} \to [0,1) \mid \textstyle\sum_n \omega(n) = 1 \text{ and } \forall n.\, \omega(n) = 0 \Rightarrow \forall m > n.\, \omega(m) = 0\}$$

Notice:

- 'shortest' in $\mathcal{D}^{\infty}_{fs<}(\mathbb{N})$ is $r\big|0\big\rangle + (1-r)\big|1\big\rangle$ for some $r \in (0,1)$
- there are inclusions $\mathcal{D}_{fs}(\underline{n}) \hookrightarrow \mathcal{D}^{\infty}_{fs<}(\mathbb{N})$, for $n > 1$

Again by finality, now of finite + infinite lists

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

Basic facts

- The composite $\mathcal{D}_{fs}(\underline{n}) \hookrightarrow \mathcal{D}^{\infty}_{fs<}(\mathbb{N}) \xrightarrow{g} (0,1)^{\infty}$ restricts to $\mathcal{D}_{fs}(\underline{n}) \to (0,1)^{+}$
- This restriction is an isomorphism. Its inverse is (finite) stick breaking:

$$(0,1)^{n-1} \xrightarrow[\cong]{sb} \mathcal{D}_{fs}(\underline{n})$$

Where we are, so far

Binomial distributions

- Urn with coloured balls is a basic model in probability theory
  - probabilities are assigned to draws, given a distribution of balls in the urn
  - The number K is used for the size of the draw (number of balls)
- Common distinctions:
  - with or without replacement. Here: **with**, so urn is constant
  - with two ("binomial") or more ("multinomial") colours. Here: **both**
- **Aim**: use stick breaking to relate binomial and multinomial distributions
  - i.e. mimick multi-colour draw as multiple draws of two colours
- Continuous version: express Dirichlet distribution in terms of multiple Beta distributions, via stickbreaking
- These results are "folklore" but hard to find explicit formulations

Binomial distributions

As channel / Kleisli map:

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

with definition, for $r \in (0,1)$,

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

Idea:

- An urn contains black & white balls (only)
- the probability of drawing a single black ball is $r \in (0,1)$
- drawn balls are inspected and returned to the urn
- consider a draw of K balls
- $bn[K](r)(i)$ is the probability that i of them are black.

Multinomial distributions, I

Intermezzo on multisets

- A multiset (bag) over a set X is a map $X \to \mathbb{N}$ with finite support
  - alternatively, it's a finite formal sum $\sum_i n_i | x_i \rangle$ of $x_i \in X$ with multiplicity $n_i \in \mathbb{N}$
  - Write $\mathcal{M}(X)$ for the set of multisets over X
- The size $\|\varphi\|$ of a multiset $\varphi = \sum_i n_i | x_i \rangle$ is the total number $\sum_i n_i$ of elements, i.e. $\|\varphi\| = \sum_x \varphi(x)$
  - write $\mathcal{M}[K](X) \hookrightarrow \mathcal{M}(X)$ for the subset of multisets of size K
- The (multinomial) coefficient (φ) is $\dfrac{\|\varphi\|!}{\prod_x \varphi(x)!}$

Multinomial distributions, II

As channel / Kleisli map:

$$\mathcal{D}_{fs}(X) \xrightarrow{mn[K]} \mathcal{D}_{fs}\Big(\mathcal{M}[K](X)\Big)$$

with definition, for $\omega \in \mathcal{D}(X)$,

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

Idea:

- X is the set of colours
- $\omega \in \mathcal{D}(X)$ the urn, as distribution of colours
- $\varphi = \sum_i n_i | x_i \rangle \in \mathcal{M}[K](X)$ is a draw of size K consisting of n_i balls of colour x_i

Note: Binomial is special case, for $X = 2 = \{0, 1\}$, since $\mathcal{D}_{fs}(2) \cong (0, 1)$ and $\mathcal{M}[K](2) \cong \{0, 1, \ldots, K\}$.

Multinomial as binomials via stick breaking

Lemma

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 $\varphi = \sum_{i<n} k_i | i \rangle \in \mathcal{M}[K](\underline{n})$,

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

We see:

- On the left of $=$: multinomial, applied to distribution $sb(\vec{r}) \in \mathcal{D}(\underline{n})$
- On the right: binomials, applied to the probabilities $r_i \in (0,1)$

Dirichlet via Beta's and stickbreaking

- Stick breaking $sb\colon (0,1)^{n-1} \xrightarrow{\cong} \mathcal{D}_{fs}(\underline{n})$ connects:
  - $n-1$ Beta distributions on $(0,1)$
  - a Dirichlet distribution on $\mathcal{D}_{fs}(\underline{n})$
  - note: Beta is binary Dirichlet
- This involves:
  - the Giry monad $\mathcal{G}$, for continuous probability
  - translation of parameters a_i

Theorem

$$\begin{aligned} &Dir(a_0, \ldots, a_{n-1}) \\ &= \mathcal{G}(sb)\Big(Beta\big(a_0, \textstyle\sum_{i>0} a_i\big) \otimes Beta\big(a_2, \textstyle\sum_{i>1} a_i\big) \otimes \cdots \\ &\qquad \cdots \otimes Beta\big(a_{n-3}, a_{n-2}+a_{n-1}\big) \otimes Beta\big(a_{n-2}, a_{n-1}\big)\Big) \end{aligned}$$

Details are in the paper.

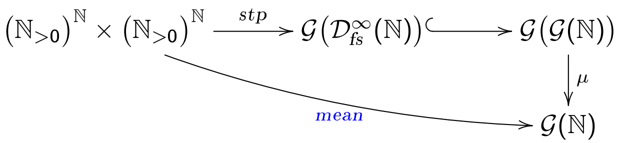
Where we are, so far

Infinite stick breaking and beta's

- For $a, b \in \mathbb{N}_{>0}$ there is $Beta(a, b) \in \mathcal{G}((0, 1))$
  - one can generalise to $a, b \in \mathbb{R}_{>0}$, but not here
- For infinite sequences (a_n), (b_n) we can form the product distribution:
$$\bigotimes_{n\in\mathbb{N}} Beta(a_n, b_n) \;\in\; \mathcal{G}\Big((0,1)^{\mathbb{N}}\Big)$$
- But now we can apply (infinite) stick breaking $(0,1)^{\mathbb{N}} \overset{\cong}{\rightrightarrows} \mathcal{D}^{\infty}_{fs}(\mathbb{N})$, giving a stick break prior:
$$sbp \;\coloneqq\; \mathcal{G}(sb)\left(\bigotimes_{n\in\mathbb{N}} Beta(a_n, b_n)\right) \;\in\; \mathcal{G}\Big(\mathcal{D}^{\infty}_{fs}(\mathbb{N})\Big)$$
- **Examples** are: Dirichlet-Poisson, Pitman-Yor
  - they are used in clustering in machine learning
  - esp. when the number of clusters is unknown and may grow

Mean of stick break prior

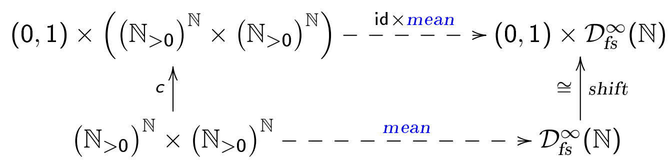
Our aim is to compute:



- This mean map actually restricts to $(\mathbb{N}_{>0})^{\mathbb{N}} \times (\mathbb{N}_{>0})^{\mathbb{N}} \longrightarrow \mathcal{D}^{\infty}_{fs}(\mathbb{N})$
- It can be computed explicitly as:
$$mean(a, b) \;=\; sb \gg\!= \left(\bigotimes_{n\in\mathbb{N}} Beta(a_n, b_n)\right) \;=\; \sum_{n\in\mathbb{N}} \frac{a_n \prod_{i<n} b_i}{\prod_{i\leq n}(a_i + b_i)}\,\big|\,n\,\big\rangle.$$
- But we can also try to exploit finality of $\mathcal{D}^{\infty}_{fs}(\mathbb{N})$.

Mean via finality

Finality works!



The coalgebra c on the left is defined as:

$$c(a, b) \;\coloneqq\; \left(\frac{a_0}{a_0 + b_0},\; a',\; b'\right) \qquad \text{where} \qquad \begin{cases} a'_n \;=\; a_{n+1} \\ b'_n \;=\; b_{n+1}. \end{cases}$$

It uses the familiar coalgebraic notation of derivative $(-)'$ of a chain.

Where we are, so far

Concluding remarks

- Stick breaking is a basic construction, with clear coalgebraic flavour — certainly in the infinite case
- It expresses "multiple" as "repeated binary"
  - in discrete probability: for multinomials and binomials
  - in continuous probability: for Dirichlet en Beta
- Stick breaking, in infinite form, emerged in stochastic process theory
  - esp. as "stick breaking priors"
  - this areay possibly harbours more coalgebraic constructions/reasoning.
- Wild idea: the stick breaking isomorphisms may be useful for combining probability with other effects, like non-determinism

Thanks for attending!