De Finetti's construction as a categorical limit

Radboud University Nijmegen

Categorical Probability and Statistics, 5-8 June 2020

Bart Jacobs, joint work with Sam Staton (Oxford)

bart@cs.ru.nl

Outline

Where we are, so far

Pólya's urn, setting

Consider an urn with b black balls and w white balls. Draw a ball, note its colour, and replace together with one extra ball of the same colour. Repeat!

- Define random variable $X_i = 0$ is the i-th draw is white; $X_i = 1$ if the draw black. This yields $X_1, X_2, X_3, \ldots$
- Let p_n be the joint distribution of $X_1, \ldots, X_n$.
- **Fact**: p_n is exchangeable, i.e. stable under permutation. E.g. $p_5(1,1,0,0,0) = p_5(0,1,0,1,0)$.

Theorem (De Finetti, binary form)

There is a unique (continuous) measure μ on $[0,1]$ such that:

$$p_n(x_1, \ldots, x_n) = \int x^t \cdot (1-x)^{n-t} \mu(\mathrm{d}x) \quad \text{where } t := \textstyle\sum_i x_i.$$

Our approach / insights / contributions

- Urn can be described as multiset
- Drawing is a coalgebra on such mulitsets
- Iterated drawing yields distribution on multiset of coloured balls
  - recall that only totals t are relevant in de Finetti
  - stability under permutation is built into multiset formalism
  - indeed, order of elements is irrelevant in a multiset — only multiplicities of elements count
- These distributions on multisets form a cone for an infinite chain
  - subtle point: cone in Kleisli category
- De Finetti's distribution μ arises because $[0, 1]$ is limit cone
- Informal idea behind "de Finetti":
  - consistent discrete distributions arise from continuous one
  - here: diagram in Kleisli category of discrete distribution monad has limit in Kleisli category of continuous distribution monad (Giry)

Where we are, so far

Multisets

- A multiset is a set in which elements may occur multiple times
- Convenient notation as formal sum $\sum_i n_i | x_i \rangle$, where $n_i \in \mathbb{N}$ is the multiplicity of element $x_i \in X$, like in a color mix (or urn):

$$3| R \rangle + 5| G \rangle + 2| B \rangle$$

- Multisets can also be described as functions with finite support:

$$\mathcal{M}(X) \coloneqq \{\varphi\colon X \to \mathbb{N} \mid \text{supp}(\varphi) \text{ is finite}\}$$

where $\text{supp}(\varphi) = \{x \in X \mid \varphi(x) \neq 0\}$.

- This is functorial: for $f\colon X \to Y$ one gets $\mathcal{M}(f)\colon \mathcal{M}(X) \to \mathcal{M}(Y)$ with:

$$\mathcal{M}(f)\Big(\sum_i n_i | x_i \rangle\Big) \coloneqq \sum_i n_i | f(x_i) \rangle.$$

- This $\mathcal{M}\colon \mathbf{Sets} \to \mathbf{Sets}$ is in fact a monad.

Multisets, continued

- Just to recall, $\mathcal{M}(X)$ is the free commutative monoid on the set X
  - addition is point-wise; empty multiset is zero-element
- We shall write $\mathcal{M}_*(X)$ for the subset of non-empty multisets:

$$\begin{aligned}\mathcal{M}_*(X) &\coloneqq \{\varphi\colon X \to \mathbb{N} \mid \mathrm{supp}(\varphi) \text{ is non-empty \& finite}\} \\ &= \{\varphi \in \mathcal{M}(X) \mid \textstyle\sum_x \varphi(x) \neq 0\}\end{aligned}$$

- For $K \in \mathbb{N}$ we write $\mathcal{M}[K](X)$ for the subset of multisets over X with K elements in total. Thus:

$$\mathcal{M}[K](X) \coloneqq \{\varphi \in \mathcal{M}(X) \mid \textstyle\sum_x \varphi(x) = K\}.$$

- **Example**: $3|R\rangle + 5|G\rangle + 2|B\rangle$ has 10 elements, i.e. is in $\mathcal{M}[10](\{R, G, B\})$

Discrete distributions

- A (finite, discrete probability) distribution is a convex combination of finitely many elements
- It's written as formal convex sum $\sum_i r_i | x_i \rangle$, where $r_i \in [0,1]$ is the multiplicity of element $x_i \in X$, where $\sum_i r_i = 1$.
  - such a distribution is also called a state
  - notice that $\mathcal{D}(2) \cong [0,1]$, where $2 = \{0,1\}$.
- Again we have monad $\mathcal{D}$: **Sets** $\to$ **Sets**
  - we call a Kleisli map $f\colon X \to \mathcal{D}(Y)$ a channel
  - notation with circle-on-arrow: $X \circ\!\!\rightarrow Y$.
- For $f\colon X \circ\!\!\rightarrow Y$ and $\omega \in \mathcal{D}(X)$ we can do state transformation $f \gg \omega$, giving a distribution on Y:

$$(f \gg \omega)(y) \coloneqq \sum_{x \in X} \omega(x) \cdot f(x)(y).$$

Binomial distributions, in the form of a channel

- Recall: for trial probability $r \in [0,1]$, the binomial distribution describes the probability of k out of K succesful trials.
- Here it is a channel $binom[K]\colon [0,1] \rightsquigarrow \{0,1,\ldots,K\}$ with:

$$binom[K](r) := \sum_{0 \le k \le K} \binom{K}{k} \cdot r^k \cdot (1-r)^{K-k} \, \big| \, k \, \big\rangle$$

- **Note**: k successes out of K can be described as a multiset $k|\,1\,\rangle + (K-k)|\,0\,\rangle$ in $\mathcal{M}[K](2)$, for $2 = \{0,1\}$.
- Indeed, there is an isomorphism:

$$\{0,1,\ldots,K\} \xrightarrow{\;\cong\;} \mathcal{M}[K](2)$$
$$k \longmapsto k|\,1\,\rangle + (K-k)|\,0\,\rangle$$

- Hence we can write $binom[K]\colon \mathcal{D}(2) \rightsquigarrow \mathcal{M}[K](2)$

  This formulation generalises easily to the multinomial case.

Multinomial distributions

- The binomial case involves a trial with two outcomes. In the multinomial case multiple outcomes are possible
- Natural formulation/generalisation as channel:

$$\mathcal{D}(X) \xrightarrow{\;mulnom[K]\;}\!\!\circ\; \mathcal{M}[K](X).$$

- With formula:

$$mulnom[K]\Big(\underbrace{\textstyle\sum_i r_i|\,x_i\,\rangle}_{\text{distribution}}\Big) \;=\; \underbrace{\sum_{k_i,\sum_i k_i = K} \underbrace{\frac{K!}{\prod_i k_i!}\cdot\prod_i r_i^{k_i}}_{\text{probability}} \,\Big|\, \underbrace{\textstyle\sum_i k_i|\,x_i\,\rangle}_{\text{multiset}}\Big\rangle}_{\text{distribution over multisets}}$$

Continuous distributions

- For a measurable space $X = (X, \Sigma_X)$ a (continuous) distribution is a countably additive map $\omega\colon \Sigma_X \to [0,1]$ with $\omega(X) = 1$.
- One writes $\mathcal{G}(X)$ for the set of such distributions
  - it forms the Giry monad
  - for a set X, as discrete measurable space: $\mathcal{D}(X) \hookrightarrow \mathcal{G}(X)$
  - this gives an inclusion of Kleisi categories
- We use continous distributions only on $[0,1]$

Where we are, so far

Drawing from an urn

- An urn with K items in total from a set X is a multiset in $\mathcal{M}[K](X)$
- Drawing-and-deleting one element yields a channel:

$$\mathcal{M}[K+1](X) \xrightarrow{\;DD\;} \mathcal{D}\Big(\mathcal{M}[K](X)\Big)$$

$$\varphi \longmapsto \sum_{x\in X} \frac{\varphi(x)}{K+1}\,\big|\,\varphi - 1|\,x\,\rangle\,\big\rangle$$

- For instance:

$$\begin{aligned} DD\Big(3|\,R\,\rangle + 5|\,G\,\rangle + 2|\,B\,\rangle\Big) &= \tfrac{3}{10}\big|\,2|\,R\,\rangle + 5|\,G\,\rangle + 2|\,B\,\rangle\,\big\rangle \\ &\quad + \tfrac{5}{10}\big|\,3|\,R\,\rangle + 4|\,G\,\rangle + 2|\,B\,\rangle\,\big\rangle \\ &\quad + \tfrac{2}{10}\big|\,3|\,R\,\rangle + 5|\,G\,\rangle + 1|\,B\,\rangle\,\big\rangle \end{aligned}$$

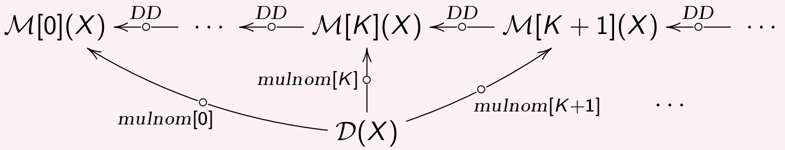
- We are interested in the infinite diagram of channels:

$$1 = \mathcal{M}[0](X) \xleftarrow{\;DD\;} \mathcal{M}[1](X) \xleftarrow{\;DD\;} \mathcal{M}[2](X) \xleftarrow{\;DD\;} \cdots$$

Multinomials form a cone

Lemma

The multinomial channels form a cone for the chain of draw-and-delete channels, as in:

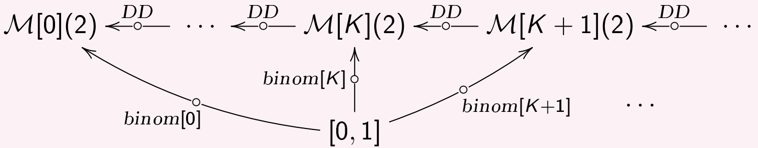


The proof is a matter of unpacking the relevant definitions: Kleisli composition, multinomials, draw-and-delete.

De Finetti's theorem, reformulated in limit form

Theorem (De Finetti, binary version)

For $X = 2$, the cone of binonomial channels forms a limit in the Kleisli category of $\mathcal{G}$, with apex $\mathcal{D}(2) \cong [0,1]$, as in:



The proof uses Hausdorff's moments theorem: a measure μ on $[0,1]$ can be obtained from a completely monotone sequence, forming its moments $\mathfrak{m}_K = \int_0^1 x^K \mu(\,\mathrm{d}x)$. A cone yields such a completely monotone sequence.

Back to the Pólya's urn example

- A Pólya urn is a multiset $b|1\rangle + w|0\rangle$ in $\mathcal{M}(2)$
  - we assume it is non-empty, that is, $b + w > 0$
- Drawing-and-recording the draws K times yields a channel:

$$\mathcal{M}_*(2) \xrightarrow{pol[K]} \mathcal{M}[K](2)$$

- These $pol[K]$ channels form a cone for the draw-and-delete diagram
- By de Finetti, this give a unique mediating map $\mathcal{M}_*(2) \to \mathcal{G}([0,1])$
  - it is the beta channel $beta\colon \mathcal{M}_*(2) \to [0,1]$
  - Explicitly, for a measurable subset $M \subseteq [0,1]$,

$$beta\Big(b|1\rangle + w|0\rangle\Big)(M) = \int_M \frac{x^b \cdot (1-x)^w}{B(b,w)}\,\mathrm{d}x$$

  - This factorisation for Pólya is well-known; what's new is that it is mediating for the limit cone of binomials

Where we are, so far

Coalgebras

- Recall that a coalgebra for an endofunctor $F\colon \mathbb{C} \to \mathbb{C}$ is a map $c\colon X \to F(X)$. A homomorphism is a map $f\colon X \to Y$ with:

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ {\scriptstyle c}\uparrow & & \uparrow{\scriptstyle d} \\ X & \xrightarrow{f} & Y \end{array}$$

- We are interested in coalgebras $c\colon X \to \mathcal{D}(2 \times X)$ for $F = \mathcal{D}(2 \times -)$.
- Call such a coalgebra exchangeable if:

$$\begin{array}{ccccc} & \overset{c}{\nearrow} & 2\times X & \xrightarrow{\mathrm{id}\otimes c} & 2\times 2\times X \\ X & & & & \cong\downarrow{\scriptstyle swap\times\mathrm{id}} \\ & \underset{c}{\searrow} & 2\times X & \xrightarrow[\mathrm{id}\otimes c]{} & 2\times 2\times X \end{array}$$

The Pólya coalgebra

The setting of the Pólya urn gives a coalgebra:

$$\mathcal{M}_*(2) \xrightarrow{\quad pol \quad} \mathcal{D}\Big(2 \times \mathcal{M}_*(2)\Big)$$

$$\Big(b|1\rangle + w|0\rangle\Big) \longmapsto \frac{b}{b+w}\Big|1,\ (b+1)|1\rangle + w|0\rangle\Big\rangle + \frac{w}{b+w}\Big|0,\ b|1\rangle + (w+1)|0\rangle\Big\rangle$$

This coalgebra is exchangeable

Iterating the coalgebra

Each coalgebra $c\colon X \rightsquigarrow 2 \times X$ gives $c_K\colon X \rightsquigarrow \mathcal{M}[K](2)$ via iteration and projection:

$$\begin{array}{ccccccccc}
X & \xrightarrow{c} & 2 \times X & \xrightarrow{\mathrm{id} \otimes c} & 2 \times 2 \times X & \longrightarrow & \cdots & \longrightarrow & 2^K \times X \\
&&&&&&&& \downarrow{\scriptstyle \pi_1} \\
&&&&&&&& 2^K \\
&&&&&&&& \downarrow{\scriptstyle \text{accumulate}} \\
&&&&&&&& \mathcal{M}[K](2)
\end{array}$$

- These maps $c_K\colon X \to \mathcal{M}[K](2)$ form a cone for the draw-and-delete chain $DD\colon \mathcal{M}[K+1](2) \to \mathcal{M}[K](2)$
- De Finetti now gives a unique mediating $\overline{c}\colon X \rightsquigarrow [0,1]$

De Finetti in terms of finality

There is a Bernouilli coalgebra:

$$\begin{array}{rcl} [0,1] & \xrightarrow{\;bern\;} & \mathcal{D}\Big(2 \times [0,1]\Big) \\ r & \longmapsto & r\big|1,\, r\big\rangle + (1-r)\big|0,\, r\big\rangle \end{array}$$

It is easy to see that it is exchangeable.

Theorem

The Bernouilli coalgebra is the final exchangeable coalgebra:

$$\begin{array}{ccc} 2 \times X & \xrightarrow{\;id \otimes \overline{c}\;} & 2 \times [0,1] \\ \uparrow c & & \uparrow bern \\ X & \xrightarrow{\;\overline{c}\;} & [0,1] \end{array}$$

Where we are, so far

Concluding remarks

- Main contribution: new, modern perspective on old, classical result of de Finetti (from 1930s)
  - could be useful in axiomatic/synthetic approaches to probability
- Accepted for *Coalgebraic Methods in Computer Science* (CMCS'20)
  - but the workshop never happened
  - the paper is available at arxiv.org/abs/2003.01964
- The interaction between multisets and distributions is essential in basic probability theory — esp. also in learning
- Channel-based approach is fruitful, also for many other topics in probability: Bayesian networks, logic, conjugate priors, etc.
- Possibilities for "de Finetti as limit" beyond the binary case:
  - via "multivariate Hausdorff moments" (Kleiber & Stoyanov, JMA'13)
  - via Polish spaces (Dahlqvist, Danos, Garnier, Concur'16)