Copy-composition for probabilistic graphical models

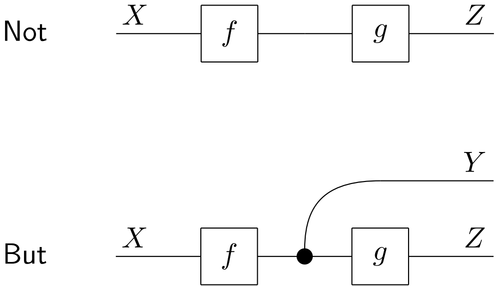
Toby St Clere Smithe

VERSES Research
&
Topos Institute

2024 June 17

Copy-composition?

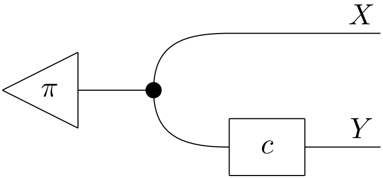
Sometimes, we want composition to "remember the intermediate object":



Copy-composition in probabilistic modelling

Often *joint distributions* are of more interest than their marginals.

Compose 'prior' $\pi : 1 \rightsquigarrow X$ and 'likelihood' $c : X \rightsquigarrow Y$ to yield



.

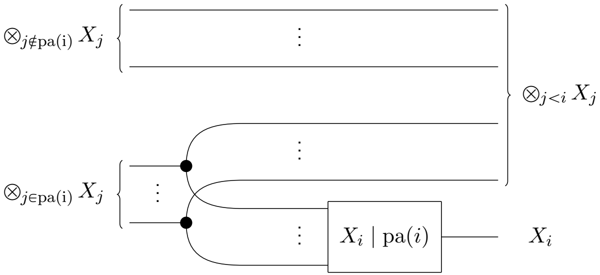
Similarly, statistical loss functions are only 'copy-compositional'.
(This was the content of my talk last year.)

Overview of this talk

Bayesian networks

A *Bayesian network* is a distribution that factorizes over a directed graph (nodes are random variables; edges represent conditional dependence).

Fong [5, Theorem 4.5]: any Bayesian network can be written as the composite of morphisms of the form

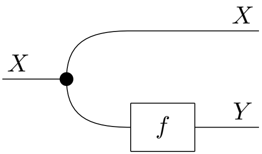


.

Note the copiers!

Pre-composing with copy yields the graph

Given a function $f : X \to Y$, its *graph* is the function



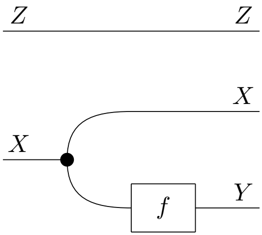
graph(f) is a section of the projection $X \times Y \to X$.

Alternatively, a 'term' of type $X \times Y$ in 'context' X.

We can tell this story in a stochastic setting.

Tensoring with identity wires is pullback

To obtain this function



pull graph(f) back along the projection $Z \times X \to X$.

Alternatively, extend its context by Z.

We can do this, too: but we'll need a (bi)fibration.

Kernels, fibrewise

Define a fibration $\mathcal{K}$ over $\mathbf{Meas}$: a "stochastic codomain fibration".
Objects of fibre $\mathcal{K}_B$ are measurable functions into B.
Morphisms $k : (E, \pi) \rightsquigarrow (E', \pi')$ are s-finite kernels[1] "fibrewise over B":

i.e., $E \overset{k}{\rightsquigarrow} E' \overset{\delta_{\pi'}}{\rightsquigarrow} B = E \overset{\delta_{\pi}}{\rightsquigarrow} B$.

Suppose $b : J \to B$ measurable. Let $p[b]$ be the pullback object:

$$\begin{array}{ccc} p[b] & \xrightarrow{\pi_E} & E \\ {\scriptstyle b^*p}\downarrow & \lrcorner & \downarrow{\scriptstyle p} \\ J & \xrightarrow[b]{} & B \end{array}$$

Then k restricts to a kernel 'fibrewise', $k[b] : p[b] \rightsquigarrow p'[b]$.

(This is Prop. 2.8 in the paper.)

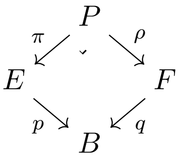
[1] An s-finite kernel $k : E \to E'$ is a function $E \times \Sigma_{E'} \to \mathbb{R}_+$, measurable in the 1st and a measure in the 2nd argument, satisfying a finiteness condition.

Bifibration structure

Substitution. $\Delta_b : \mathcal{K}_B \to \mathcal{K}_J$ acts by pullback on objects, and by restriction on kernels, mapping k to $k[b]$.

Dependent sum. $\Sigma_b : \mathcal{K}_J \to \mathcal{K}_B$ acts by post-composition on objects, mapping $E \xrightarrow{p} J$ to $E \xrightarrow{p} J \xrightarrow{b} B$; and as identity on kernels. $\Sigma \dashv \Delta$.

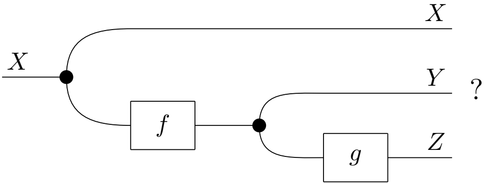
Beck-Chevalley. Given pullback in $\mathbf{Meas}$



we have $\Sigma_\rho \, \Delta_\pi \;\cong\; \Delta_q \, \Sigma_p$, naturally. (This follows from the situation in $\mathbf{Meas}^{\to}$.)

Copy-composition is pull-push

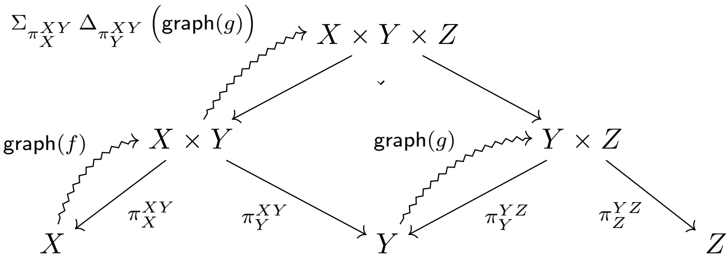
Can composing $\mathrm{graph}(f) : X \to X \times Y$ and $\mathrm{graph}(g) : Y \to Y \times Z$ yield



Yes, by 'pull-push'! The diagram represents a section $X \to X \times Y \times Z$.

It may be obtained as $\Sigma_{\pi^{XY}_X} \Delta_{\pi^{XY}_Y} \left(\mathrm{graph}(g)\right) \circ \mathrm{graph}(f)$ ...

Copy-composition is pull-push



We have decorated spans with sections (of their left legs)!

A decorated span (double) category

We can formalize this via a double Grothendieck construction, following Patterson [7] and Cruttwell, Lambert, Pronk, and Szyld [2].

Given a (B-C) bifibration $\pi : \mathcal{E} \to \mathcal{B}$ with a section ι, we obtain a double fibration $\mathbb{S}$ over $\mathbb{S}\mathbf{pan}(\mathcal{B})_{pb}$.[2]

Spans in $\mathcal{B}$ are decorated by left-leg sections in $\mathcal{E}$, via ι.

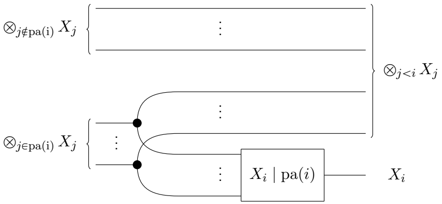
Horizontal composition is pull-push on the decorations.

Beck-Chevalley needed for associativity.

Details in the paper!

[2] 2-cells are Cartesian natural transformations, needed for horizontal composition of squares.

Composing Bayesian networks



sfKrn embeds into $\mathbb{S}$ (lax, horiz.), mapping a kernel to its graph.

Let σ be the image of $X_i \mid \mathrm{pa}(i)$ under this embedding,
and let p be the projection $\prod_{j<i} X_j \to \prod_{j \in \mathrm{pa}(i)} X_j$.

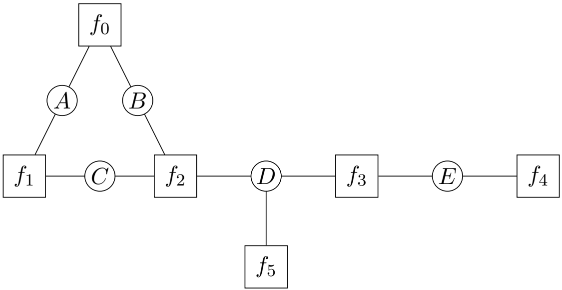
Then Fong's morphism is $\sigma_{\leqslant i} = \Delta_p \, \sigma$.

More generally: Bayesian networks with dependent types...

From directed to undirected models

Factor graphs

In the 'bipartite' style of Wainwright and Jordan [8, §2.1.3]:



$$f(a, b, c, d, e) = f_0(a, b)\, f_1(a, c)\, f_2(b, c, d)\, f_3(d, e)\, f_4(e)\, f_5(d)$$

$$f : A \times B \times C \times D \times E \to \mathbb{R}_+$$

Each factor may itself contain a factor graph...

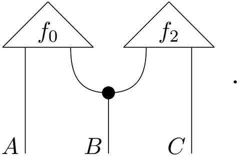
Factors, costates, and copy-composition

In $\mathbf{sfKrn}$, measurable functions $X \to \mathbb{R}_+$ are kernels into 1.

This is a common pattern: factors are 'predicates'.

In copy-discard cat.s[3], these are *costates*: morphisms into the monoidal unit.

And we can copy-compose them via



A classic case of decorated cospans ...

[3] Such as: categories of modules; weakly [6] / partial [4] Markov categories; partial effectuses [1] ...

A decorated cospan double category, $\mathbb{FG}$

We'll use the full double categorical machinery this time.

Factors' domains are finite tensors of objects.
So decorate finite sets X with type information, 'interfaces' $\chi : X \to \mathcal{C}_0$.

A decorated cospan double category, $\mathbb{FG}$

We'll use the full double categorical machinery this time.

Factors' domains are finite tensors of objects.
So decorate finite sets X with type information, 'interfaces' $\chi : X \to \mathcal{C}_0$.

A cospan encodes those objects which are exposed for composition.

A decorated cospan double category, $\mathbb{FG}$

We'll use the full double categorical machinery this time.

Factors' domains are finite tensors of objects.
So decorate finite sets X with type information, 'interfaces' $\chi : X \to \mathcal{C}_0$.

A cospan encodes those objects which are exposed for composition.

Vertical morphisms in $\mathbb{C}\mathbf{ospan}(\mathbf{FinSet})$ are functions;
decorate these with comonoid homomorphisms (so they can be copied!).

A decorated cospan double category, $\mathbb{FG}$

We'll use the full double categorical machinery this time.

Factors' domains are finite tensors of objects.
So decorate finite sets X with type information, 'interfaces' $\chi : X \to \mathcal{C}_0$.

A cospan encodes those objects which are exposed for composition.

Vertical morphisms in $\mathbb{C}\mathbf{ospan}(\mathbf{FinSet})$ are functions;
decorate these with comonoid homomorphisms (so they can be copied!).

Decorate apices X with interfaces χ *and* factors $f : \chi^{\otimes} \rightsquigarrow I$.
Formally, this is again a Grothendieck construction (factors over interfaces).

A decorated cospan double category, $\mathbb{FG}$

We'll use the full double categorical machinery this time.

Factors' domains are finite tensors of objects.
So decorate finite sets X with type information, 'interfaces' $\chi : X \to \mathcal{C}_0$.

A cospan encodes those objects which are exposed for composition.

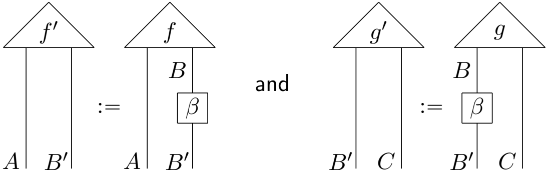
Vertical morphisms in $\mathbb{C}\mathbf{ospan}(\mathbf{FinSet})$ are functions;
decorate these with comonoid homomorphisms (so they can be copied!).

Decorate apices X with interfaces χ *and* factors $f : \chi^{\otimes} \rightsquigarrow I$.
Formally, this is again a Grothendieck construction (factors over interfaces).

Decorations on squares obtain accordingly:
briefly, morphisms in $\mathcal{C}$ that are deterministic over exposed interfaces.

Morphisms of factors: deterministic if exposed

Suppose β transforms $f \to f'$ and $g \to g'$ as in

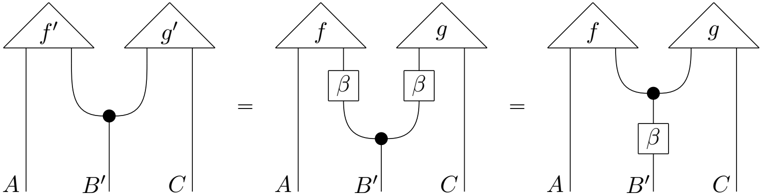


and we want to compose along β, over cospans like

$$\{A\} \to \{A, B'\} \leftarrow \{B'\} \to \{B', C\} \leftarrow \{C\} \ .$$

Morphisms of factors: deterministic if exposed

Naturality of copy-composition requires



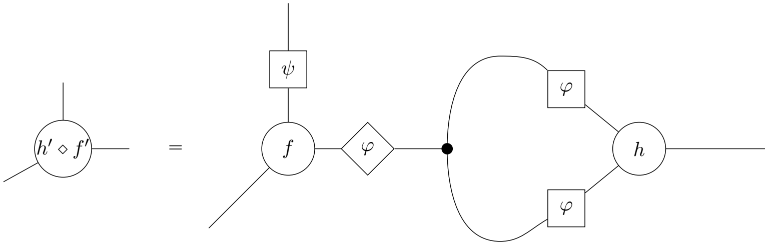
hence β must be a homomorphism.

But we *can have unexposed* non-deterministic transformations,
e.g. marginalization (compose with a state).

Double-categorical undirected wiring diagrams

Absent the vertical morphisms, this is an undirected wiring diagram algebra.

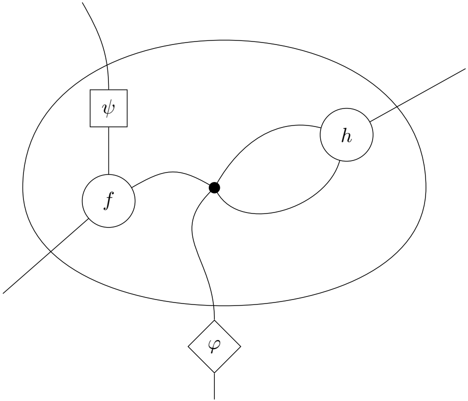
Suggests a graphical calculus, now with 2-cells.



'Spiders' represent loci of composition.
f, h transform to f', h'. φ deterministic, ψ not.

Double-categorical undirected wiring diagrams

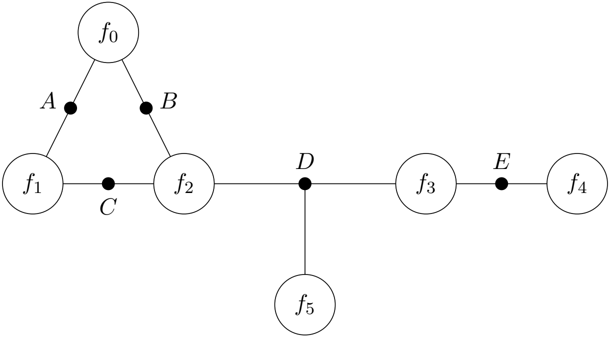
Can pull deterministic transformations outside the bubble



(exposing the corresponding locus)

Double-categorical undirected wiring diagrams

The factor graph we started with:



Very similar to those of de Vries and Friston [3] (& colleagues).

But this calculus not yet formalized ...

Review & some future directions

We saw two cases of copy-composition in probabilistic modelling:
one directed, one undirected.

Both were formalized by a Grothendieck construction,
associating models ('terms') to interfaces ('types').

Might hope to find directed models amongst undirected ones,
as functions are amongst relations.
But, for that, these constructions aren't quite right.

And it would be nice to formalize the 2-d calculus for UWDs.

Finally: is $\mathbb{FG}$ useful for compositional belief propagation?

Thanks!

References I

1. Cho, K., Jacobs, B., Westerbaan, B., and Westerbaan, A.: “An Introduction to Effectus Theory”
2. Cruttwell, G., Lambert, M., Pronk, D., and Szyld, M.: Double Fibrations. (2022). arXiv: 2205.15240 [math]. http://arxiv.org/abs/2205.15240 (visited on 03/20/2023). preprint
3. De Vries, B., and Friston, K.J.: A Factor Graph Description of Deep Temporal Active Inference. Frontiers in Computational Neuroscience **11** (2017). DOI: 10.3389/fncom.2017.00095
4. Di Lavore, E., and Román, M.: Evidential Decision Theory via Partial Markov Categories. (2023). arXiv: 2301.12989 [cs, math]. http://arxiv.org/abs/2301.12989 (visited on 04/20/2023). preprint
5. Fong, B.: Causal Theories: A Categorical Perspective on Bayesian Networks. University of Oxford (2013-01-26, 2013)

References II

6. Fritz, T., Gadducci, F., Perrone, P., and Trotta, D.: Weakly Markov Categories and Weakly Affine Monads. LIPIcs, Volume 270, CALCO 2023 **270**, 16:1–16:17 (2023). DOI: 10.4230/LIPIcs.CALCO.2023.16. arXiv: 2303.14049 [cs, math]. http://arxiv.org/abs/2303.14049 (visited on 06/10/2024)

7. Patterson, E.: Structured and Decorated Cospans from the Viewpoint of Double Category Theory. (2023). arXiv: 2304.00447 [math]. http://arxiv.org/abs/2304.00447 (visited on 04/20/2023). preprint

8. Wainwright, M.J., and Jordan, M.I.: Graphical Models, Exponential Families, and Variational Inference. Foundations and Trends® in Machine Learning **1**(1–2), 1–305 (2007). DOI: 10.1561/2200000001. http://dx.doi.org/10.1561/2200000001