Compositional inference for Bayesian networks and causality

Bart Jacobs, **Márk Széles**, Dario Stein

Computing with Markov Categories Workshop
Tallinn, February 26, 2025

Radboud Universiteit

Outline

- Discrete probability subdistributions
- Copy-discard-compare categories
- Normalisation structures in copy-discard categories
- Disintegration and Bayesian inversion
- A compositional calculus of conditioning

Discrete probability subdistributions

Definition

A **discrete probability subdistribution** on a set X is a function $\omega : X \to [0, 1]$ such that

1. $\mathrm{supp}(\omega) = \{x \in X : \omega(x) \neq 0\}$ is a finite set.
2. $\|\omega\| = \sum_{x \in X} \omega(x) \leq 1$.

Discrete probability subdistributions

Definition

A **discrete probability subdistribution** on a set X is a function $\omega : X \to [0,1]$ such that

1. $\mathrm{supp}(\omega) = \{x \in X : \omega(x) \neq 0\}$ is a finite set.
2. $\|\omega\| = \sum_{x \in X} \omega(x) \leq 1$.

- Write subdistributions as formal sums $\omega = \sum_{x \in X} \omega(x)|x\rangle$

Discrete probability subdistributions

Definition

A **discrete probability subdistribution** on a set X is a function $\omega : X \to [0,1]$ such that

1. $\mathrm{supp}(\omega) = \{x \in X : \omega(x) \neq 0\}$ is a finite set.
2. $\|\omega\| = \sum_{x \in X} \omega(x) \leq 1$.

- Write subdistributions as formal sums $\omega = \sum_{x \in X} \omega(x)|x\rangle$
- There is a monad $\mathcal{D}_{\leq} : \mathbf{Sets} \to \mathbf{Sets}$.

Discrete probability subdistributions

Definition

A **discrete probability subdistribution** on a set X is a function $\omega : X \to [0,1]$ such that

1. $\text{supp}(\omega) = \{x \in X : \omega(x) \neq 0\}$ is a finite set.
2. $\|\omega\| = \sum_{x \in X} \omega(x) \leq 1$.

- Write subdistributions as formal sums $\omega = \sum_{x \in X} \omega(x)|x\rangle$
- There is a monad $\mathcal{D}_{\leq} : \mathbf{Sets} \to \mathbf{Sets}$.
- Write $f : X \rightsquigarrow Y$ for a Kleisli-map $f : X \to \mathcal{D}_{\leq}(Y)$.

Discrete probability subdistributions

Definition

A **discrete probability subdistribution** on a set X is a function $\omega : X \to [0,1]$ such that

1. $\mathrm{supp}(\omega) = \{x \in X : \omega(x) \neq 0\}$ is a finite set.
2. $\|\omega\| = \sum_{x \in X} \omega(x) \leq 1$.

- Write subdistributions as formal sums $\omega = \sum_{x \in X} \omega(x)|x\rangle$
- There is a monad $\mathcal{D}_{\leq} : \mathbf{Sets} \to \mathbf{Sets}$.
- Write $f : X \rightsquigarrow Y$ for a Kleisli-map $f : X \to \mathcal{D}_{\leq}(Y)$.
- Write $f(y|x)$ for $f(x)(y) \in [0,1]$.

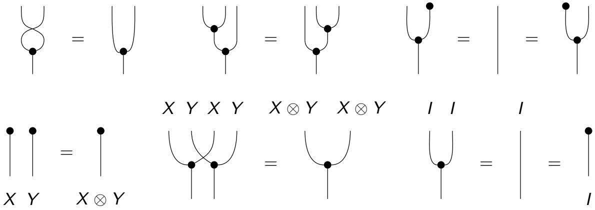
Copy-discard-compare (CDC) categories

Definition

A **copy-discard-compare category** is a symmetric monoidal category $(C, \otimes, I)$ with **copier** $\Delta_X : X \to X \otimes X$, **discard** $d_X : X \to I$, and **comparator** $\nabla_X : X \otimes X \to X$ maps such that...

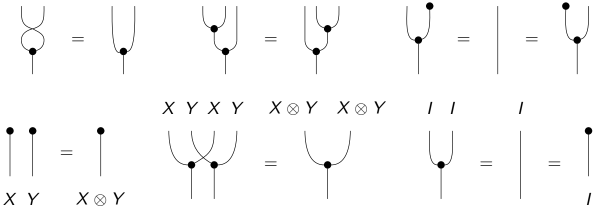
The axioms of copy-discard-compare (CDC) categories

- Copy and discard:

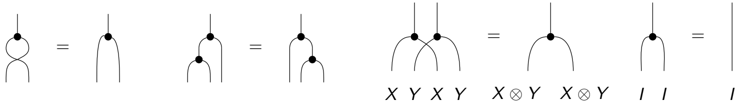


The axioms of copy-discard-compare (CDC) categories

- Copy and discard:

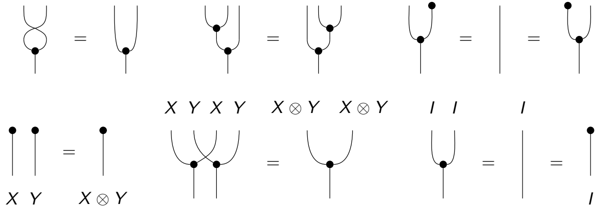


- Compare:

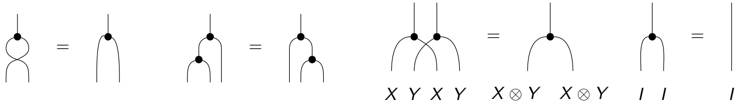


The axioms of copy-discard-compare (CDC) categories

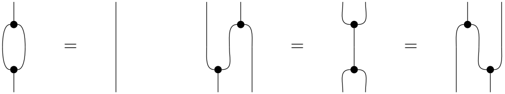
- Copy and discard:



- Compare:

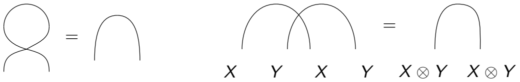


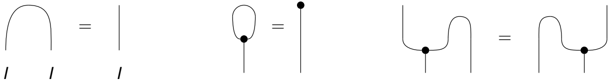
- Copy-compare interaction:



Alternative definitions of CDC-categories

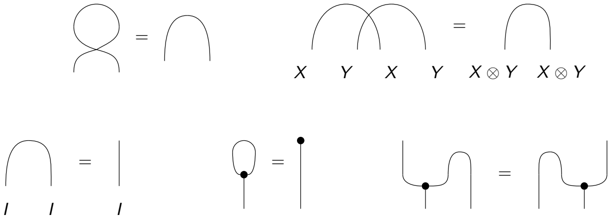
- CDC-categories can also be defined in terms of caps $\cap_X : X \otimes X \to I$.





Alternative definitions of CDC-categories

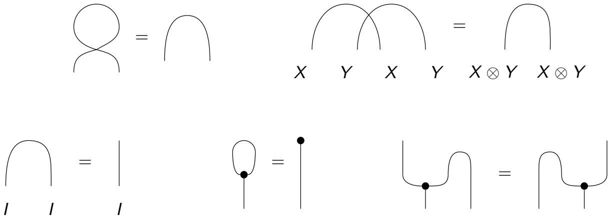
- CDC-categories can also be defined in terms of caps $\cap_X : X \otimes X \to I$.



- Caps and comparators are inter-definable.

Alternative definitions of CDC-categories

- CDC-categories can also be defined in terms of caps $\cap_X : X \otimes X \to I$.



- Caps and comparators are inter-definable.

- There is a third, equivalent, definition via 'least' disintegrations of copier and identity maps.

Normalisation

- Non-zero subdistributions $\omega \in \mathcal{D}_{\leq}(X)$ can by normalised:

$$\mathrm{nrm}(\omega) = \sum_{x \in X} \frac{\omega(x)}{\|\omega\|} \big| x \big\rangle$$

Normalisation

- Non-zero subdistributions $\omega \in \mathcal{D}_{\leq}(X)$ can by normalised:

$$\mathrm{nrm}(\omega) = \sum_{x \in X} \frac{\omega(x)}{\|\omega\|} \big| x \big\rangle$$

- Subprobability kernels $f : X \rightsquigarrow Y$ can be normalised pointwise:

$$\mathrm{nrm}(f)(x) = \begin{cases} \mathrm{nrm}(f(x)) & \text{if } \|f(x)\| \neq 0 \\ \mathbf{0} & \text{otherwise} \end{cases}$$

Normalisation

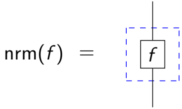
- Non-zero subdistributions $\omega \in \mathcal{D}_{\leq}(X)$ can by normalised:

$$\mathrm{nrm}(\omega) = \sum_{x \in X} \frac{\omega(x)}{\|\omega\|} \big| x \big\rangle$$

- Subprobability kernels $f : X \multimap Y$ can be normalised pointwise:

$$\mathrm{nrm}(f)(x) = \begin{cases} \mathrm{nrm}(f(x)) & \text{if } \|f(x)\| \neq 0 \\ \mathbf{0} & \text{otherwise} \end{cases}$$

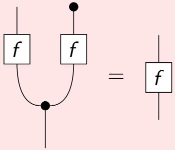
- Graphically, we write



Self-normalising maps

Definition

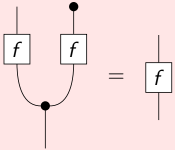
We call a map $f : X \to Y$ **self-normalising** if



Self-normalising maps

Definition

We call a map $f : X \to Y$ **self-normalising** if



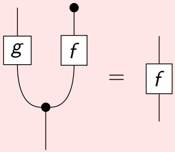
For a subchannel $f : X \multimap Y$ in $\mathcal{K}l(\mathcal{D}_{\leq})$ this translates to

$$\forall x \in X. \|f(x)\| \in \{0, 1\}$$

The ‘normalised by’ relation

Definition

A map $g : X \to Y$ **normalises** $f : X \to Y$ if



In this case, we write $f \preceq g$.

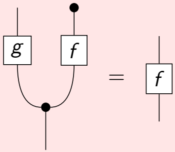
For subchannels $f, g : X \rightsquigarrow Y$ in $\mathcal{K}l(\mathcal{D}_{\leq})$ this translates to

$$\forall x \in X. f(x) \neq \mathbf{0} \implies g(x) = \mathrm{nrm}(f(x))$$

The 'normalised by' relation

Definition

A map $g : X \to Y$ **normalises** $f : X \to Y$ if



In this case, we write $f \preceq g$.

For subchannels $f, g : X \multimap Y$ in $\mathcal{K}l(\mathcal{D}_{\leq})$ this translates to

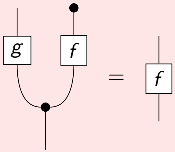
$$\forall x \in X. f(x) \neq \mathbf{0} \implies g(x) = \mathrm{nrm}(f(x))$$

- The relation $\preceq$ is a partial order on self-normalising maps.

The 'normalised by' relation

Definition

A map $g : X \to Y$ **normalises** $f : X \to Y$ if



In this case, we write $f \preceq g$.

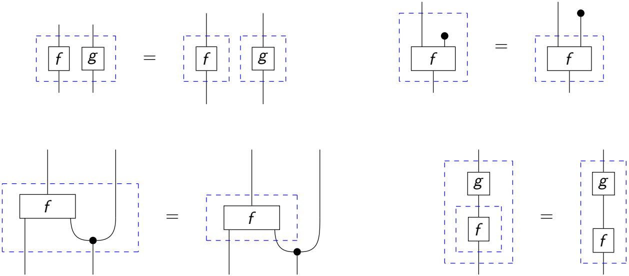
For subchannels $f, g : X \multimap Y$ in $\mathcal{K}l(\mathcal{D}_{\leq})$ this translates to

$$\forall x \in X. f(x) \neq \mathbf{0} \Longrightarrow g(x) = \mathrm{nrm}(f(x))$$

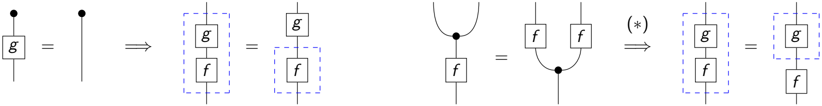
- The relation $\preceq$ is a partial order on self-normalising maps.
- The dashed box should select the least normalisation.

Properties of normalisation

- Equational properties:



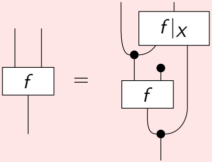
- Implicational properties:



Disintegration

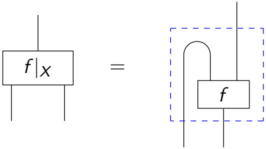
Definition

If $f : Z \to X \otimes Y$, then a **disintegration** of f is a map $f|_X : X \otimes Z \to Y$ that satisfies



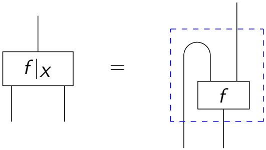
Building your own disintegrations

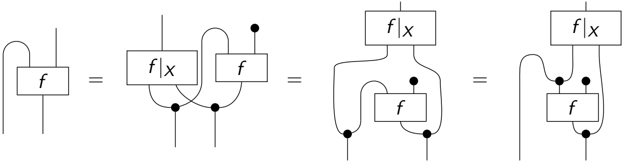
Try to define a disintegration by



Building your own disintegrations

Try to define a disintegration by

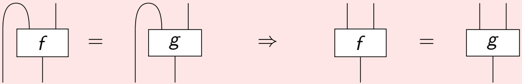




Building your own disintegrations

Definition

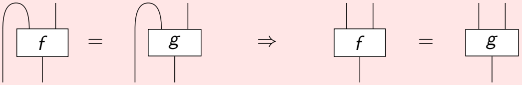
A CDC-category has **cancellative caps** if the following implication holds for all $f, g : X \to Y \otimes Z$:



Building your own disintegrations

Definition

A CDC-category has **cancellative caps** if the following implication holds for all $f, g : X \to Y \otimes Z$:



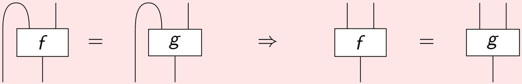
Proposition (Lorenz and Tull, 2023)

In a CDC-category with normalisation and cancellative caps, every map has a disintegration.

Building your own disintegrations

Definition

A CDC-category has **cancellative caps** if the following implication holds for all $f, g : X \to Y \otimes Z$:



Proposition (Lorenz and Tull, 2023)

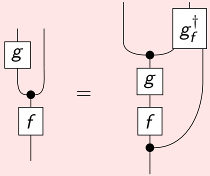
In a CDC-category with normalisation and cancellative caps, every map has a disintegration.

This does not recover disintegration in BorelStoch$_{\leq 1}$!

Bayesian inversions

Definition

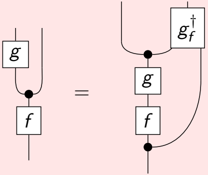
Let $g : Y \to Z$, $f : X \to Y$. A **Bayesian inversion** of g with respect to f is a map $g_f^\dagger : Z \otimes X \to Y$ that satisfies



Bayesian inversions

Definition

Let $g : Y \to Z$, $f : X \to Y$. A **Bayesian inversion** of g with respect to f is a map $g_f^\dagger : Z \otimes X \to Y$ that satisfies

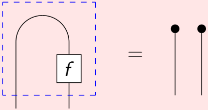


Bayesian inversions exist $\iff$ Disintegrations exist

Broad support

Definition

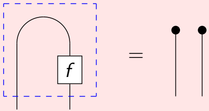
Let $\mathbb{C}$ be a CDC-category with cancellative caps. A map $f : X \to Y$ in $\mathbb{C}$ has **broad support** if



Broad support

Definition

Let $\mathbb{C}$ be a CDC-category with cancellative caps. A map $f : X \to Y$ in $\mathbb{C}$ has **broad support** if

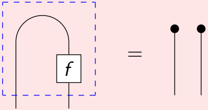


- For $f : X \rightsquigarrow Y$ this means $f(x)(y) > 0$ for all $x \in X$, $y \in Y$.

Broad support

Definition

Let $\mathbb{C}$ be a CDC-category with cancellative caps. A map $f : X \to Y$ in $\mathbb{C}$ has **broad support** if

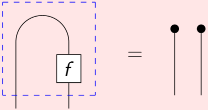


- For $f : X \multimap Y$ this means $f(x)(y) > 0$ for all $x \in X$, $y \in Y$.
- Maps with broad support are closed under many operations: composition, tensor, marginalisation, normalisation, disintegration.

Broad support

Definition

Let $\mathbb{C}$ be a CDC-category with cancellative caps. A map $f : X \to Y$ in $\mathbb{C}$ has **broad support** if

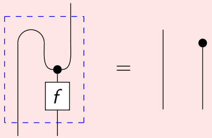


- For $f : X \rightsquigarrow Y$ this means $f(x)(y) > 0$ for all $x \in X$, $y \in Y$.
- Maps with broad support are closed under many operations: composition, tensor, marginalisation, normalisation, disintegration.
- If a self-normalising f has broad support, then f is total.

Disappearence of dashed boxes

Proposition

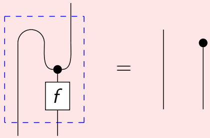
If $f : X \to Y$ has broad support, then



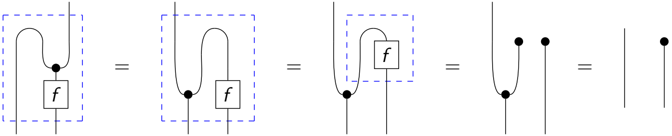
Disappearence of dashed boxes

Proposition

If $f : X \to Y$ has broad support, then



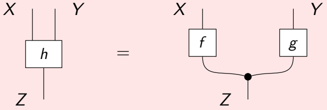
Proof:



Example: deriving conditional independence

Definition

Let $h : X \to Y \otimes Z$. We say that, in the context of h, Y is **conditionally independent** of Z given X, if h can be factorised as



In this case, we write $Y \amalg Z \mid X$.

Example: deriving conditional independence

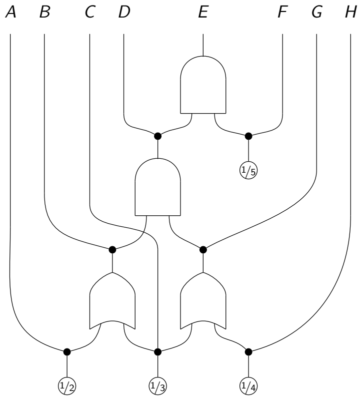
Prove that $A \coprod B \mid Z$ in the following program:

```
Z <- flip(1/2)
X <- flip(1/2)
Y <- flip(1/2)
A = X || Z
B = Z || Y
return (Z, X, Y, A, B)
```

Example: deriving conditional independence

Example: inference in a Bayesian network

Compute how conditioning on B effects E in the following fault tree:



Example: inference in a Bayesian network

Conclusion and outlook

- We developed a powerful compositional calculus for computing disintegrations in discrete probabilistic models.
- We recognised the role of Bayesian inversion and the notion of broad support.
- Many more examples in the paper:
    - Applications to probabilistic programming.
    - We give an elegant derivation of the 'front-door-adjustment' formula, a known example from the causal reasoning literature.
    - An example of counterfactual reasoning.

Conclusion and outlook

- We developed a powerful compositional calculus for computing disintegrations in discrete probabilistic models.
- We recognised the role of Bayesian inversion and the notion of broad support.
- Many more examples in the paper:
  - Applications to probabilistic programming.
  - We give an elegant derivation of the 'front-door-adjustment' formula, a known example from the causal reasoning literature.
  - An example of counterfactual reasoning.
- Question: how to generalise to continuous probability?
- Outlook: implementation in a string diagram rewrite tool.