ACT 2023 1 August 2023

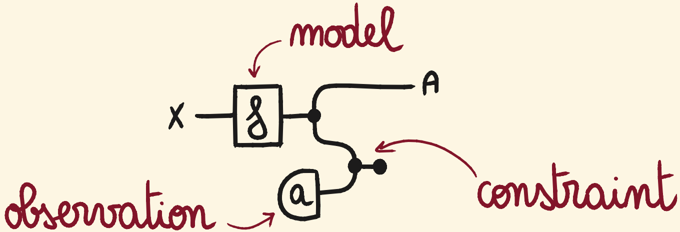
PARTIAL MARKOV CATEGORIES

Elena Di Lavore and Mario Román

Tallinn University of Technology

MOTIVATION

- Find the algebraic structure to express belief updates.
- Markov categories express probabilistic processes.



Updating a model on an observation means restricting the model to scenarios that are compatible with this observation.

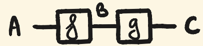
OUTLINE

- Copy-discard categories
- Markov categories
- Cartesian restriction categories
- Partial Markov categories

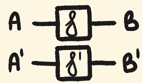
STRING DIAGRAMS

$\mathcal{C}$ symmetric monoidal category

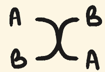
- composition $f;g : A \to C$
  for $f : A \to B$, $g : B \to C$ in $\mathcal{C}$

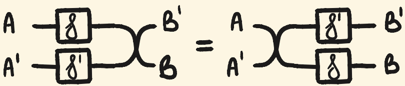


- monoidal product $f \otimes f' : A \otimes A' \to B \otimes B'$
  for $f : A \to B$, $f' : A' \to B'$ in $\mathcal{C}$



- symmetry $\sigma_{A,B} : A \otimes B \to B \otimes A$





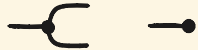
(naturality)

EXAMPLES

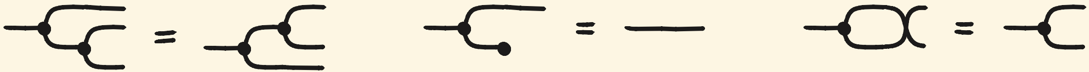
- $(\mathbf{Set}, \times, \{*\})$: sets and functions
  $f : A \to B$ is a function
- $(\mathbf{Par}, \times, \{*\})$: sets and partial functions
  $f : A \nrightarrow B$ is a function $f : A \to B + 1$
- $(\mathcal{K}\ell(\mathcal{D}), \times, \{*\})$: sets and stochastic functions
  $f : A \nrightarrow B$ is a function $f : A \to \mathcal{D}(B)$
- $(\mathbf{Rel}, \times, \{*\})$: sets and relations
  $f : A \nrightarrow B$ is a function $f : A \to P(B)$
- $(\mathcal{K}\ell(\mathcal{D}_{\leq 1}), \times, \{*\})$: sets and partial stochastic functions
  $f : A \nrightarrow B$ is a function $f : A \to \mathcal{D}(B + 1)$

COPY-DISCARD CATEGORIES

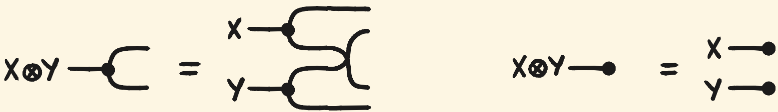
A copy-discard category is a symmetric monoidal category where every object is a uniform cocommutative comonoid.



COCOMMUTATIVE COMONOID



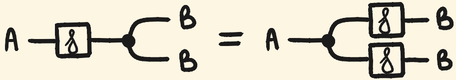
UNIFORMITY



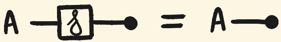
[Corradini & Gadducci 1999]

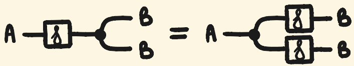
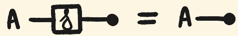
DETERMINISTIC & TOTAL MAPS

Deterministic maps can be copied.



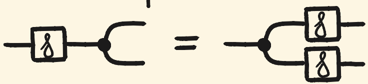
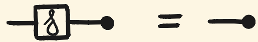
Total maps can be discarded.



EXAMPLES		
$(\mathbf{Set}, \times, \{*\})$	✓	✓
$(\mathbf{Par}, \times, \{*\})$	✓	✗
$(\mathcal{K}\ell(\mathcal{D}), \times, \{*\})$	✗	✓
$(\mathbf{Rel}, \times, \{*\})$	✗	✗
$(\mathcal{K}\ell(\mathcal{D}_{\leq 1}), \times, \{*\})$	✗	✗

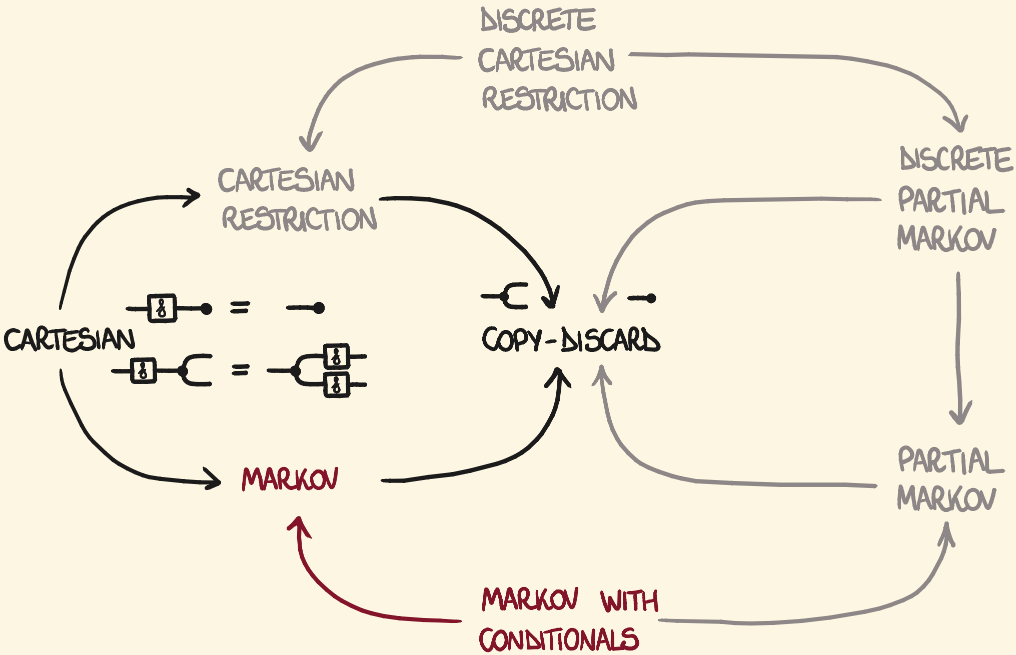
FOX'S THEOREM

A copy-discard category is cartesian if and only if all morphisms are deterministic and total,

 and  for all f.

[Fox 1976]

OUTLINE


DISCRETE CARTESIAN RESTRICTION
CARTESIAN RESTRICTION
DISCRETE PARTIAL MARKOV
CARTESIAN
COPY-DISCARD
MARKOV
PARTIAL MARKOV
MARKOV WITH CONDITIONALS

PROBABILISTIC PROCESSES

Markov categories express probabilistic processes, for example

- throwing a coin 

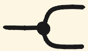
- tomorrow's weather given today's clouds 

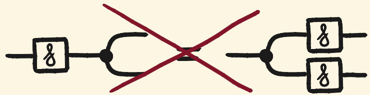
- developing cancer given smoking habits 

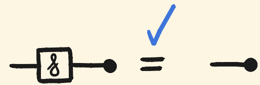
MARKOV CATEGORIES & CONDITIONALS

A Markov category with conditionals is a copy-discard category with conditionals where all morphisms are total.

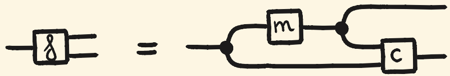
COPY-DISCARD STRUCTURE







CONDITIONALS



[Fritz 2020]

FINITARY DISTRIBUTIONS

A <u>finitary distribution</u> $\sigma \in \mathcal{D}(A)$ is a function $\sigma : A \to [0,1]$ such that

- its support, $\mathrm{supp}(\sigma) := \{a \in A \mid \sigma(a) > 0\}$, is finite, and
- its total probability mass is 1, $\sum_{a \in A} \sigma(a) = 1$.

A morphism $X - \boxed{f} - A$ in $\mathrm{Kl}\mathcal{D}$ is a function $X \to \mathcal{D}(A)$

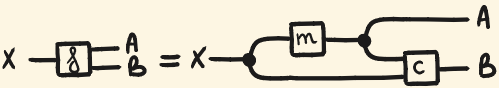
$$f(a \mid x) = \text{“probability of } a \text{ given } x\text{”}$$

composition is

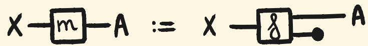
$$X - \boxed{f} - \boxed{g} - B \; (b \mid x) := \sum_{a \in A} f(a \mid x) \cdot g(b \mid a)$$

CONDITIONALS

$\mathrm{Kl}\mathcal{D}$ has conditionals.



$$m(a|x) := \sum_{b \in B} f(a,b|x)$$



$$c(b|a,x) := \begin{cases} \dfrac{f(a,b|x)}{m(a|x)} & \text{if } m(a|x) \neq 0 \\ \sigma(b) & \text{if } m(a|x) = 0 \end{cases}$$

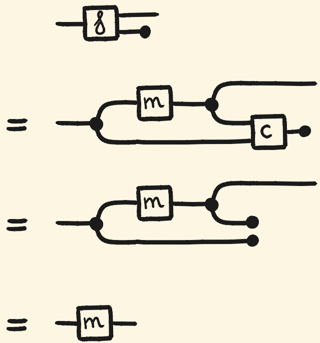
any distribution on B

$\rightsquigarrow$ conditionals are not unique and they cannot be

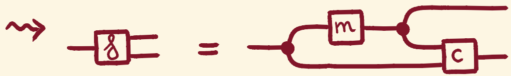
MARGINALS IN MARKOV CATEGORIES

Marginals in Markov categories are as expected :

PROOF



conditionals :



⇝ totality

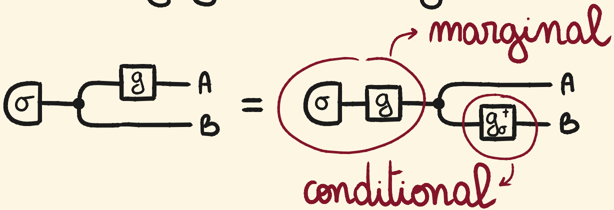
□

BAYES INVERSION

The Bayes inversion of a channel $g : B \to A$ with respect to a distribution $\sigma : I \to B$ is classically defined as

$$g^{\dagger}_{\sigma}(b|a) := \frac{g(a|b)\,\sigma(b)}{\sum_{b' \in B} g(a|b')\,\sigma(b')}$$

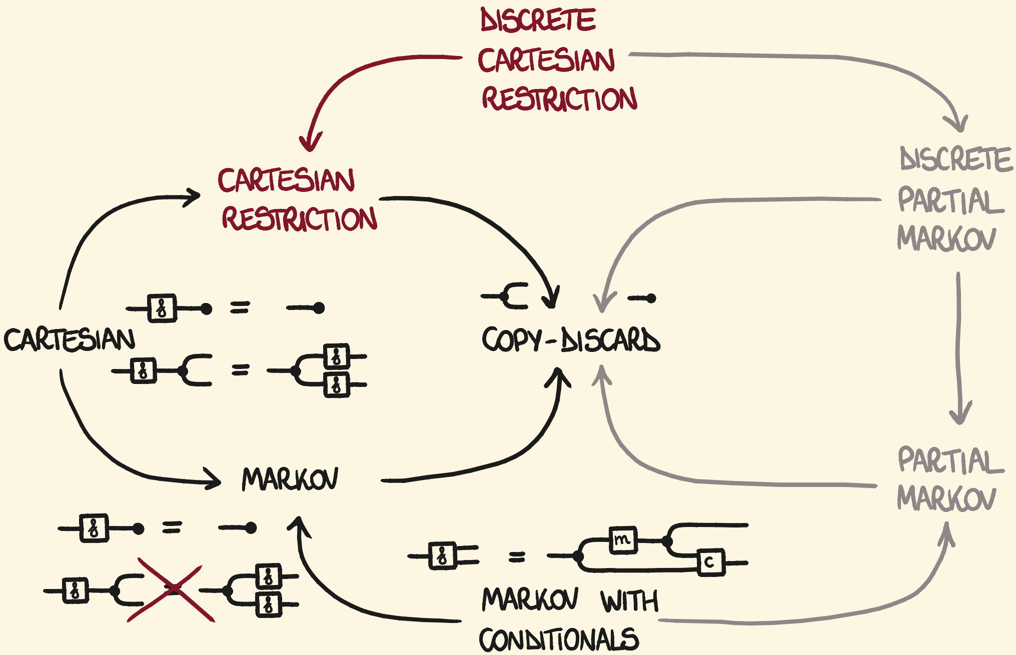
In a Markov category, it is a $g^{\dagger}_{\sigma} : A \to B$ such that



Bayes inversions are instances of conditionals.

[Cho & Jacobs 2019]

OUTLINE

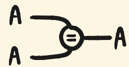

DISCRETE CARTESIAN RESTRICTION
CARTESIAN RESTRICTION
DISCRETE PARTIAL MARKOV
CARTESIAN
COPY-DISCARD
MARKOV
PARTIAL MARKOV
m
c
MARKOV WITH CONDITIONALS

PARTIAL PROCESSES

Cartesian restriction categories express partial computations, for example

- computing $\frac{1}{x}$

- checking equality

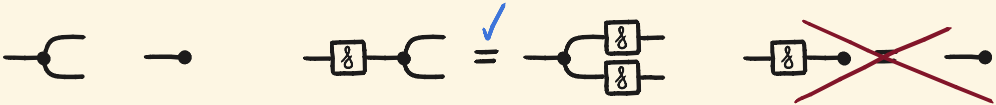


- non-terminating computations

CARTESIAN RESTRICTION CATEGORIES

A cartesian restriction category is a copy-discard category where all morphisms are deterministic.

COPY - DISCARD STRUCTURE



[Cockett & Lack 2003, 2007]

PARTIAL FUNCTIONS

$(\mathsf{Par}, \times, \{*\})$ is a cartesian restriction category

- objects are sets $A, B, C, \ldots$
- morphisms are partial functions $f : A \rightharpoonup B$, $g : B \rightharpoonup C, \ldots$
  i.e. functions $f : A \to B+1$, $g : B \to C+1, \ldots$
- composition is

$$f ; g\,(a) := \begin{cases} g(f(a)) & \text{if } f(a) \neq \bot \\ \bot & \text{if } f(a) = \bot \end{cases}$$

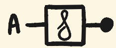
- monoidal product is

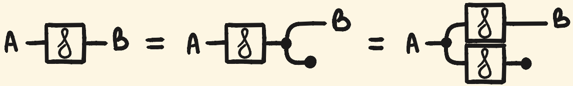
$$f \times f'\,(a, a') := \begin{cases} (f(a), f'(a')) & \text{if } f(a) \neq \bot \text{ and } f'(a') \neq \bot \\ \bot & \text{if } f(a) = \bot \text{ or } f'(a') = \bot \end{cases}$$

PREDICATES, DOMAINS, RESTRICTIONS

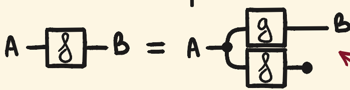
Morphisms $q : A \to 1$ in Par are <u>predicates</u>.

 $(a) = \begin{cases} * & \text{if } a \text{ satisfies } q \\ \perp & \text{if } a \text{ does not satisfy } q \end{cases}$

The <u>domain</u> of  is the predicate .



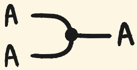
The <u>restriction preorder</u> on morphisms is

$f \leq g$ iff 

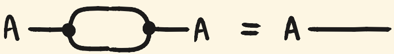
g restricted to the domain of f

EQUALITY CHECK

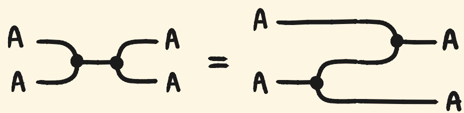
Par has equality checks.

 $(a, a') := \begin{cases} a & \text{if } a = a' \\ \perp & \text{if } a \neq a' \end{cases}$

Equality checks interact with the comonoid structure.

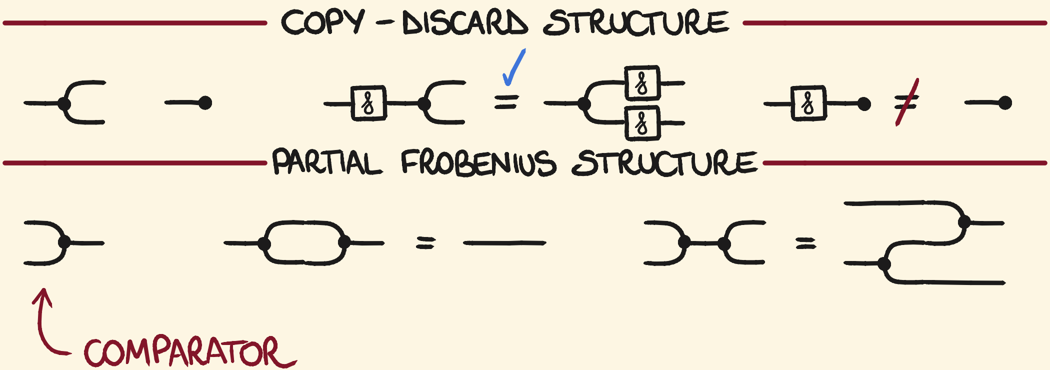


and



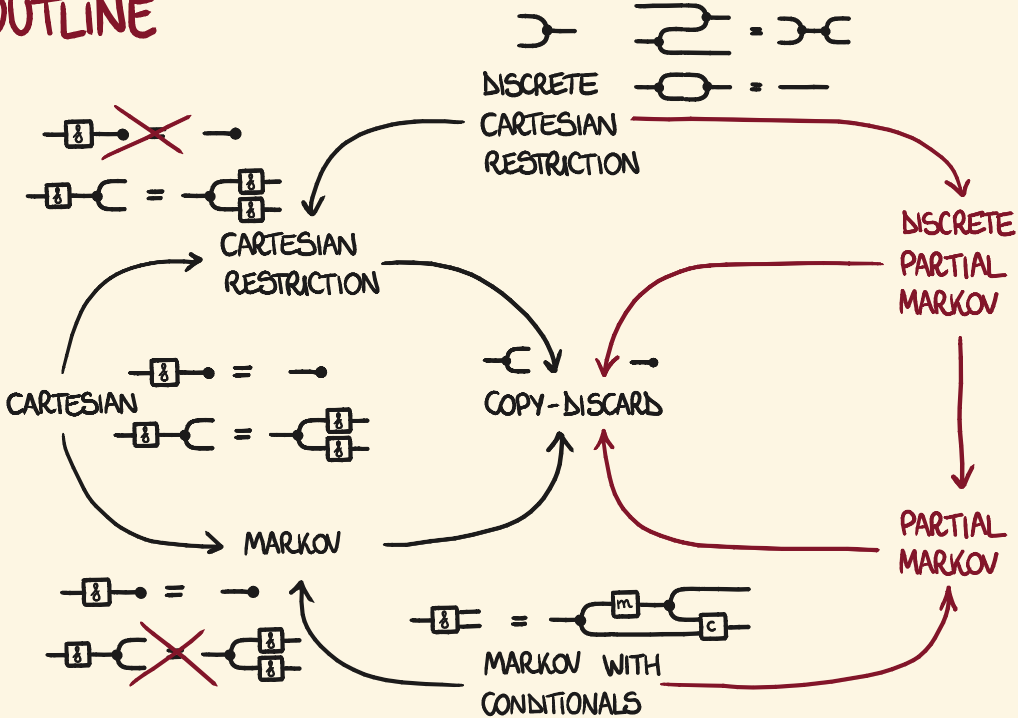
CONSTRAINTS VIA PARTIAL FROBENIUS

A discrete cartesian restriction category is a copy-discard category with comparators where all morphisms are deterministic.



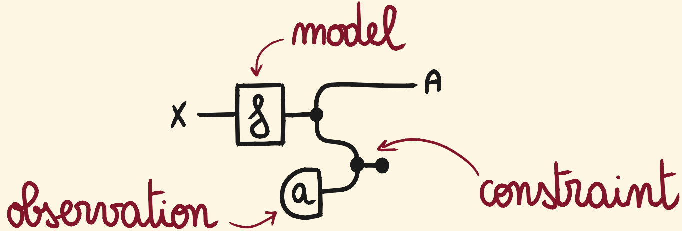
[Cockett, Guo & Hofstra 2012, Di Liberti, Loregian, Nester & Sobociński 2020]

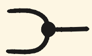
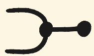
OUTLINE


DISCRETE CARTESIAN RESTRICTION
CARTESIAN RESTRICTION
DISCRETE PARTIAL MARKOV
CARTESIAN
COPY-DISCARD
MARKOV
PARTIAL MARKOV
MARKOV WITH CONDITIONALS
m
c

PARTIALITY FOR OBSERVATIONS

Updating a model on an observation means restricting the model to scenarios that are compatible with this observation.



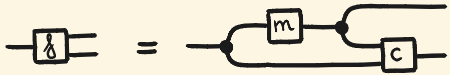
constraints  cannot be total computations because  ≠  .

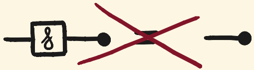
OVERVIEW

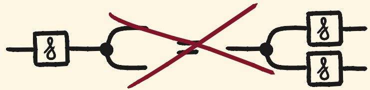
Combine Markov and cartesian restriction categories to express partial stochastic processes.

cartesian restriction

Markov with conditionals

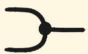


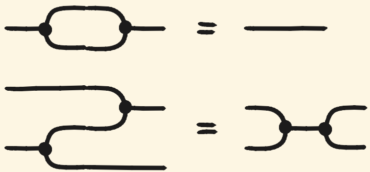




Add the discrete structure to express equality checking.

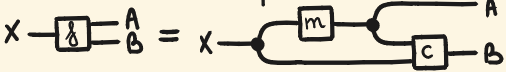
discrete cartesian restriction





DROPPING TOTALITY

We want to keep the nice marginals of Markov categories.



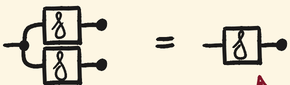
Should we ask conditionals to be total? ✗ NO

⇝ too strong: total conditionals fail to exist in $Kl(\mathcal{D}_{\leq 1})$.

Can we ask conditionals to be quasi-total? ✓ YES

⇝ sweet spot: quasi-total conditionals usually exist and give nice marginals.

QUASI-TOTAL MORPHISM (in a copy-discard category)



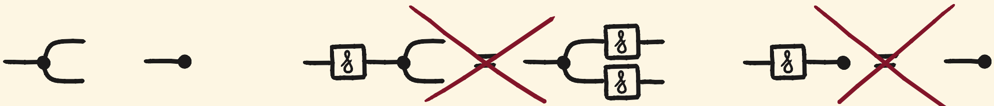
⇝ failure is deterministic

domain of definition

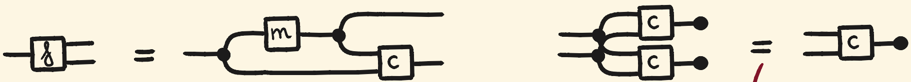
PARTIAL MARKOV CATEGORIES

A partial Markov category is a copy-discard category with quasi-total conditionals.

COPY-DISCARD STRUCTURE



CONDITIONALS



quasi-totality

[EDL & Román, 2023]

SUBDISTRIBUTIONS

A <u>subdistribution</u> σ on A is a distribution on $A+1$:

$\sigma \in \mathcal{D}_{\leq 1}(A)$ is a function $\sigma : A \to [0,1]$ such that

- its support, $\mathrm{supp}(\sigma) := \{a \in A \mid \sigma(a) > 0\}$, is finite, and
- its total probability mass is at most 1, $\sum_{a \in A} \sigma(a) \leq 1$.

A morphism $X - \boxed{f} - A$ in $\mathrm{Kl}\mathcal{D}_{\leq 1}$ is a function $X \to \mathcal{D}_{\leq 1}(A)$

$f(a|x)$ = "probability of a given x"

$f(\perp|x)$ = "probability of failure"

Composition is

$$X - \boxed{f} - \boxed{g} - B \; (b|x) := \sum_{a \in A} f(a|x) \cdot g(b|a)$$

$$X - \boxed{f} - \boxed{g} - B \; (\perp|x) := \sum_{a \in A} f(a|x) \cdot g(\perp|a) + f(\perp|x)$$

EXAMPLES : PARTIAL STOCHASTIC PROCESSES

A <u>partial stochastic process</u> is a stochastic process that may fail.

↳ Maybe monad

a Markov category with conditionals

Partial stochastic processes form a partial Markov category.

PROPOSITION

$\begin{cases} \mathcal{C} \text{ Markov category with conditionals and coproducts} \\ \text{Some ugly technical conditions} \end{cases}$

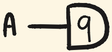
$\Rightarrow \mathrm{Kl}(\cdot + 1)$ is a partial Markov category.

EXAMPLES

- $\mathrm{Kl}(\mathcal{D}(\cdot + 1))$ ⇝ finitary subdistributions
- $\mathrm{Kl}(\mathcal{G}iry_B(\cdot + 1))$ ⇝ subdistributions on standard Borel spaces

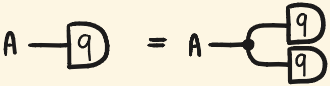
PREDICATES & DOMAINS

Morphisms $q : A \to 1$ in $\mathrm{Kl}(\mathcal{D}(\bullet + 1))$ are 'fuzzy' predicates.



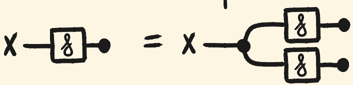
$(* \mid a)$ $\rightsquigarrow$ probability of a being true

Deterministic predicates are classical predicates.



$\Rightarrow$ q is a classical predicate

Quasi-total morphisms have a domain.



$\rightsquigarrow$ domain of f

probability of failure of f

CONDITIONALS IN SUBDISTRIBUTIONS

A quasi-total morphism $g : X \to B$ is a function $g : X \to \mathcal{D}B + 1$.

The marginal of $f : X \to A \otimes B$ is

$$X - \boxed{m} - A\,(a|x) = X - \boxed{f} \overset{A}{=\!\bullet}\,(a|x) = \sum_{b \in B} f(a, b|x)$$

$$X - \boxed{m} - A\,(\perp|x) = X - \boxed{f} \overset{A}{=\!\bullet}\,(\perp|x) = f(\perp|x)$$

A conditional of f is:

$$\begin{matrix}A\\X\end{matrix} \boxed{c} - B\,(b|a,x) = \begin{cases} \dfrac{f(a,b|x)}{m(a|x)} & m(a|x) \neq 0 \\ 0 & m(a|x) = 0 \end{cases}$$

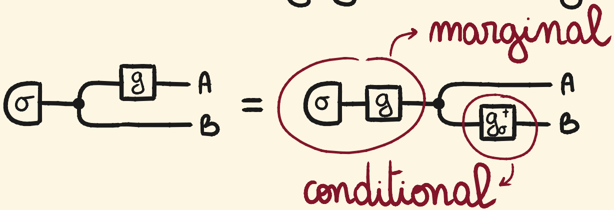
$$\begin{matrix}A\\X\end{matrix} \boxed{c} - B\,(\perp|a,x) = \begin{cases} 0 & m(a|x) \neq 0 \\ 1 & m(a|x) = 0 \end{cases}$$

BAYES INVERSION

The Bayes inversion of a channel $g : B \to A$ with respect to a distribution $\sigma : I \to B$ is classically defined as

$$g^{\dagger}_{\sigma}(b|a) := \frac{g(a|b)\,\sigma(b)}{\sum_{b' \in B} g(a|b')\,\sigma(b')}$$

In a partial Markov category, it is a $g^{\dagger}_{\sigma} : A \to B$ such that



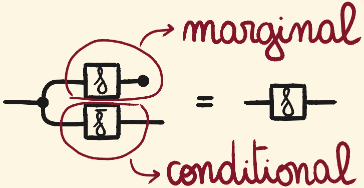
Bayes inversions are instances of quasi-total conditionals.

NORMALISATION

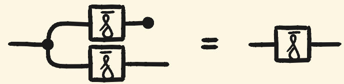
The normalisation of a partial channel $f : X \to A$ is classically defined as

$$\bar{f}(a|x) := \frac{f(x|a)}{1 - f(\perp|a)}$$

In a partial Markov category, it is a $\bar{f} : X \to A$ such that



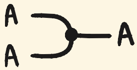
and



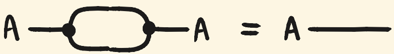
Normalisations are instances of quasi-total conditionals.

EQUALITY CHECK

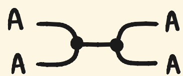
$\mathrm{Kl}\mathcal{D}_{\leq 1}$ has equality checks.

 $(a, a') := \begin{cases} \delta_a & \text{if } a = a' \\ \delta_\perp & \text{if } a \neq a' \end{cases}$

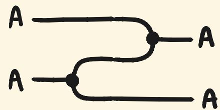
Equality checks interact with the comonoid structure.



and



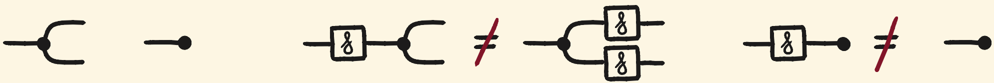
=



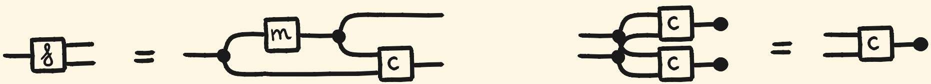
DISCRETE PARTIAL MARKOV CATEGORIES

A discrete partial Markov category is a copy-discard category with quasi-total conditionals and comparators.

COPY - DISCARD STRUCTURE



CONDITIONALS

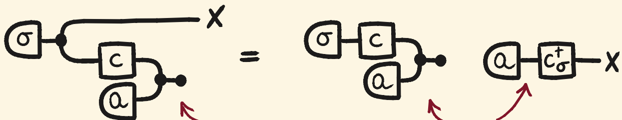


PARTIAL FROBENIUS STRUCTURE

[EDL & Román, 2023]

SYNTHETIC BAYES THEOREM

A deterministic observation $a : I \to A$ from a prior $\sigma : I \to X$ through a channel $c : X \to A$ determines an update proportional to the Bayes inversion $c^{\dagger}_{\sigma}$ evaluated on a.

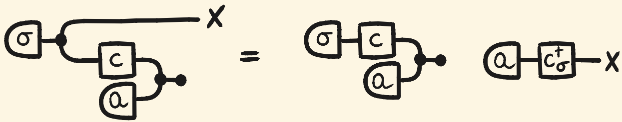


$$P(X{=}x \mid A{=}a) = \frac{P(A{=}a \mid X{=}x)\cdot P(X{=}x)}{\sum_{y\in X} P(A{=}a \mid X{=}y)\cdot P(X{=}y)}$$

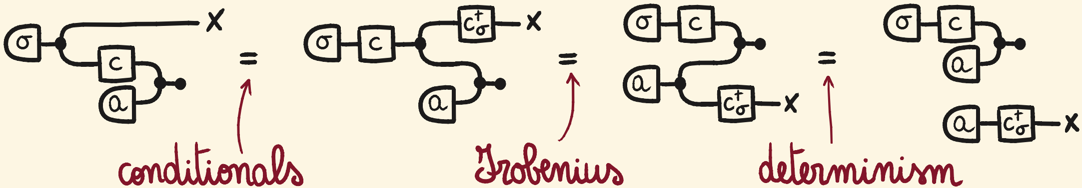
classical formula
for Bayes theorem

SYNTHETIC BAYES THEOREM

A deterministic observation $a : I \to A$ from a prior $\sigma : I \to X$ through a channel $c : X \to A$ determines an update proportional to the Bayes inversion $c^\dagger_\sigma$ evaluated on a.

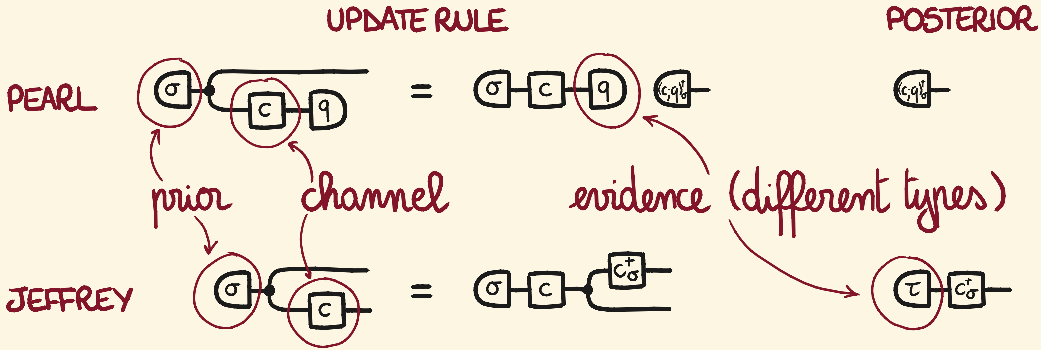


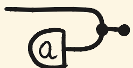
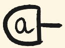
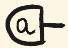
PROOF



$\square$

PEARL'S VS JEFFREY'S UPDATES

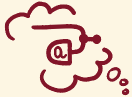


Pearl's update on  coincides with Jeffrey's update on , whenever  is deterministic.

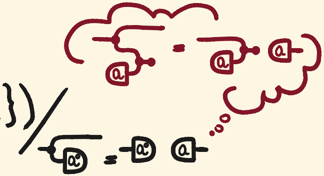
[Jacobs 2019]

PROCESSES WITH EXACT OBSERVATIONS

We can add exact observations to any Markov category:



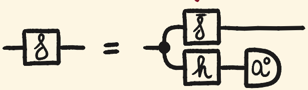
exOb($\mathscr{C}$) = ($\mathscr{C}$ + {A — a° | a — A deterministic})/



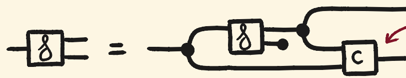
embeds faithfully into ($\mathscr{C}$ +) / partial Frobenius

Conditionals and normalisations are computed in $\mathscr{C}$

normalisation of f



conditional of f

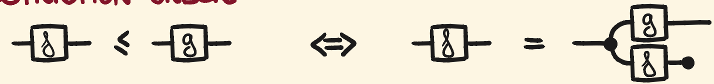


$\Rightarrow$ exOb($\mathscr{C}$) is a partial Markov category.

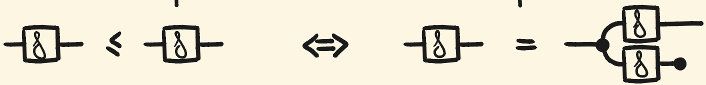
MINIMAL CONDITIONALS

$\mathcal{C}$ partial Markov category

RESTRICTION ORDER



→ this is a partial order on quasi-total morphisms



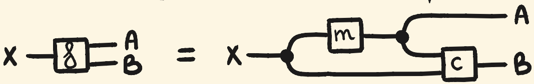
(reflexivity) (quasi-totality)

MINIMAL CONDITIONALS

$\mathcal{C}$ has minimal conditionals if, for every , the partial order on its quasi-total conditionals has a minimal element.

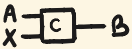
MINIMAL CONDITIONALS IN SUBDISTRIBUTIONS

A quasi-total morphism $g : X \to B$ is a function $g : X \to \mathcal{D}B + 1$.



The conditionals of f are:

$$(b \mid a, x) = \begin{cases} \dfrac{f(a, b \mid x)}{m(a \mid x)} & m(a \mid x) \neq 0 \\ \sigma(b) & m(a \mid x) = 0 \end{cases}$$



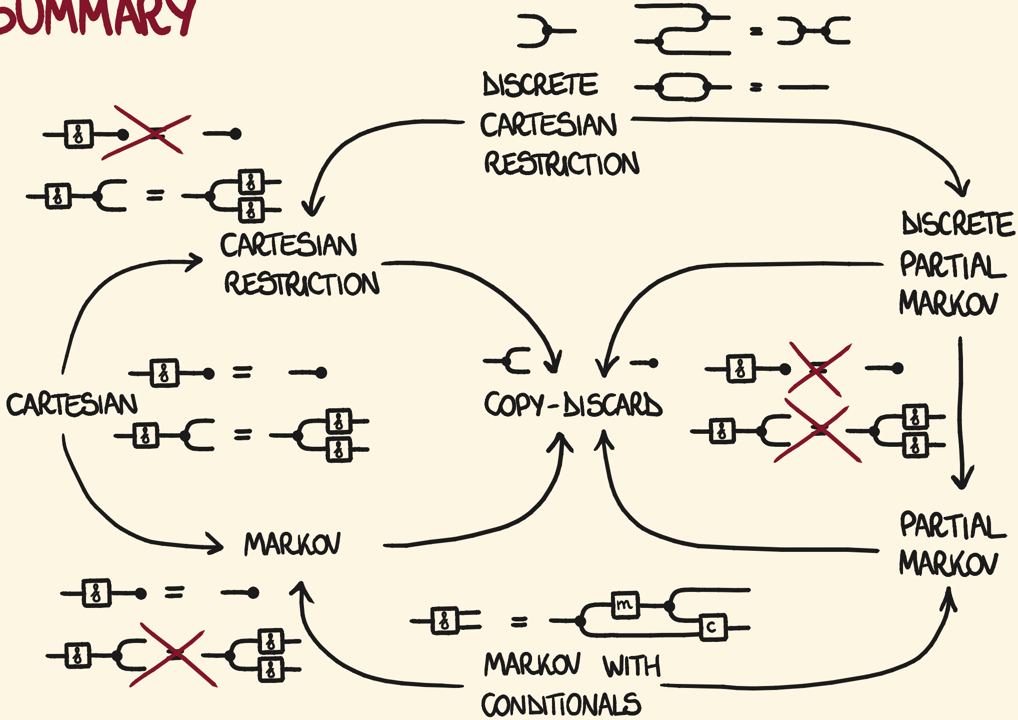
$$(\perp \mid a, x) = \begin{cases} 0 & m(a \mid x) \neq 0 \\ \sigma(\perp) & m(a \mid x) = 0 \end{cases}$$

for some $\sigma \in \mathcal{D}(B) + 1$.

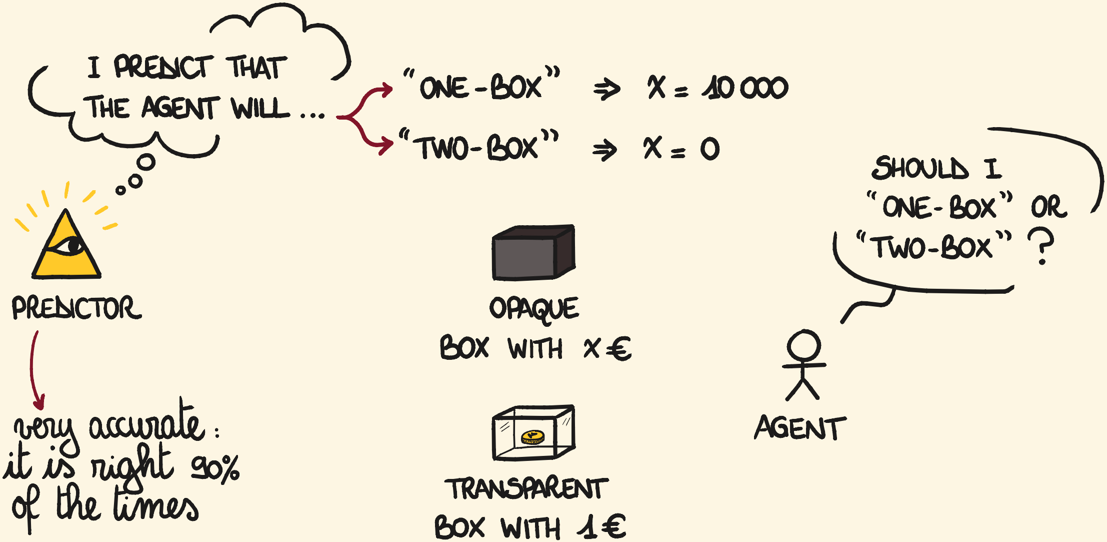
$\Rightarrow$ The minimal choice for σ is $\begin{cases} \sigma(b) = 0 & \text{for } b \in B \\ \sigma(\perp) = 1 \end{cases}$.

$\rightsquigarrow$ fail on unexpected observations

SUMMARY


DISCRETE CARTESIAN RESTRICTION
CARTESIAN RESTRICTION
DISCRETE PARTIAL MARKOV
CARTESIAN
COPY-DISCARD
PARTIAL MARKOV
MARKOV
MARKOV WITH CONDITIONALS
m
c

NEWCOMB'S PROBLEM



CAUSAL DECISION THEORY

Causal decision theory answers :

"Which action would cause the best-case scenario ?"

Whatever the predictor did,
I get 1€ extra if I two-box
⇒ I will two-box

BEST !

PREDICTOR \ AGENT	ONE-BOX	TWO-BOX
ONE-BOX	10 000 €	10 001€
TWO-BOX	0 €	1 €

EVIDENTIAL DECISION THEORY

<u>Evidential</u> decision theory answers :

"Which action would be <u>evidence</u> for the best-case scenario ?"

My action is evidence for the prediction :
if I one-box I expect 10 000 €,
if I two-box I expect 1 € .
⇒ I will <u>one-box</u>

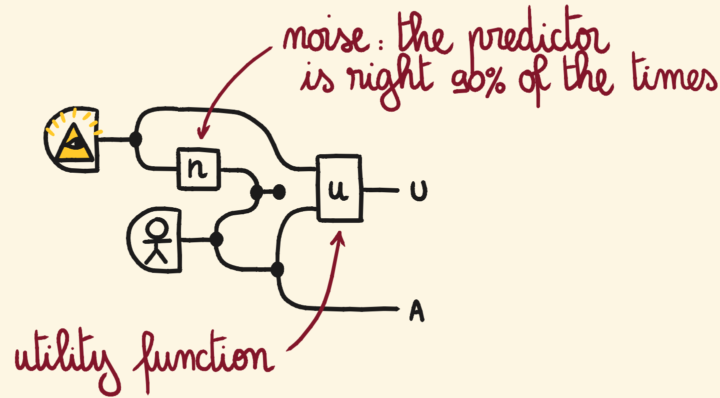
PREDICTOR \ AGENT	ONE-BOX	TWO-BOX
ONE-BOX	10 000 €	10 001 €
TWO-BOX	0 €	1 €

MOST LIKELY

EVIDENTIAL VS CAUSAL DECISION THEORY

	 CAUSAL DECISION THEORIST	 EVIDENTIAL DECISION THEORIST
EXPECTED UTILITY	0.9×1 € $+ 0.1 \times 10\,001$ € $= 1\,001$ €	$0.9 \times 10\,000$ € $+ 0.1 \times 0$ € $= 9\,000$ €

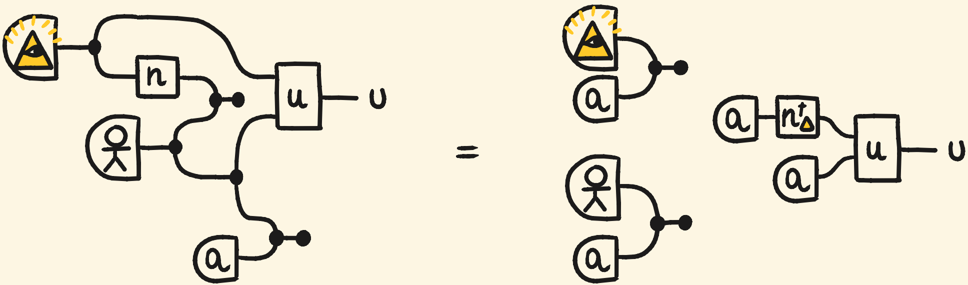
NEWCOMB'S PROBLEM CATEGORICALLY



AGENT / PREDICTOR	ONE-BOX	TWO-BOX
ONE-BOX	10 000 €	10 001 €
TWO-BOX	0 €	1 €

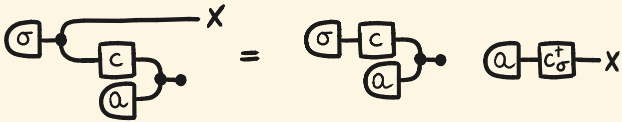
SOLVING NEWCOMB'S PROBLEM

Evidential decision theory asks:
"Which action would be evidence for the best-case scenario?"
i.e. "Which action maximises the average of the state below?"



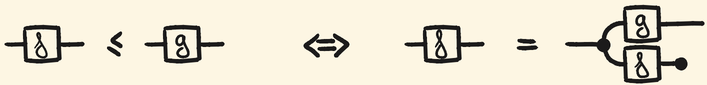
CONCLUSIONS & FUTURE WORK

- Partial Markov categories are for updating on observations.



Synthetic Bayes theorem

- Is there a bicategorical structure hiding?



Restriction preorder