PARTIAL MARKOV CATEGORIES

Mario Román

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF OXFORD

TOPOS COLLOQUIUM, 30 OCTOBER 2024

Elena Di Lavore
UNIVERSITÀ DI PISA

Bart Jacobs
RADBOUD UNIVERSITY NIJMEGEN

- 🗎 **Evidential Decision Theory via Partial Markov Categories.**
  Elena Di Lavore and Mario Román.
  Logic in Computer Science, 2023.
- 🗎 **A Simple Formal Language for Probabilistic Decision Theory.**
  Elena Di Lavore, Bart Jacobs, and Mario Román.
  arXiv Preprint. 2024.

Partially supported by COST–EuroProofNet.

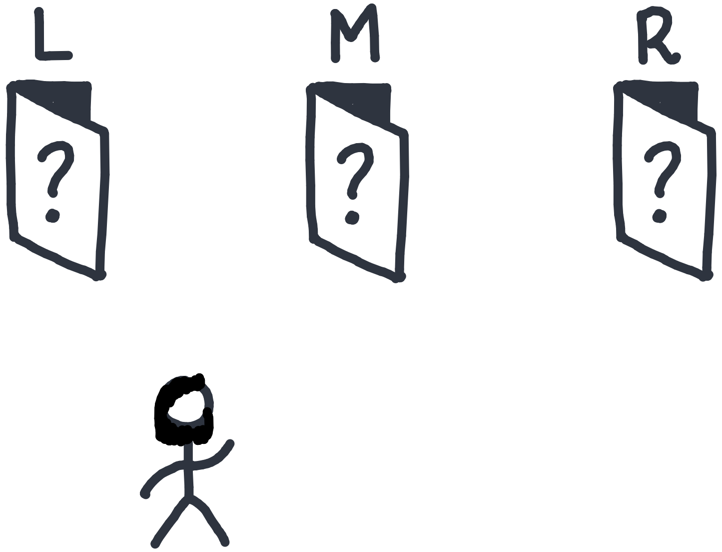
MOTIVATION

- Do we really know how to solve discrete probability problems?
  - ↳ Given disease prevalence, specificity, and sensitivity, what is the posterior after a multiset of test results? (📄 B. Jacobs, 2023)

- Formal language for decision problems: make assumptions explicit.

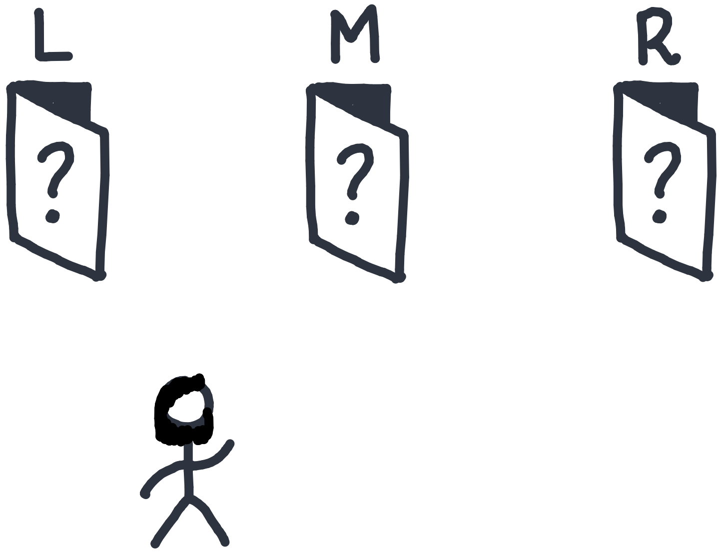
- Synthetic theory of probability, normalization, and Bayes' update.
  - ↳ What are postulates for Bayesian updating?

MONTY-HALL PROBLEM



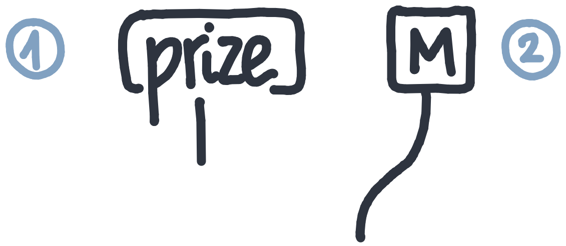
MONTY-HALL PROBLEM

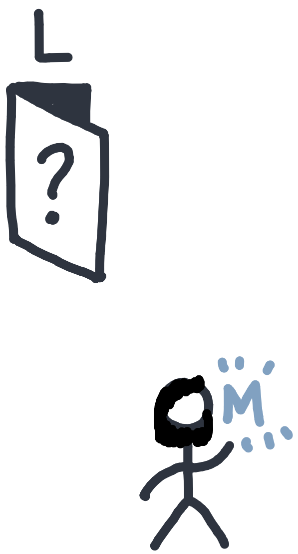
① $\boxed{\text{prize}}$



① $\frac{1}{3}|L\rangle + \frac{1}{3}|M\rangle + \frac{1}{3}|R\rangle$

MONTY-HALL PROBLEM

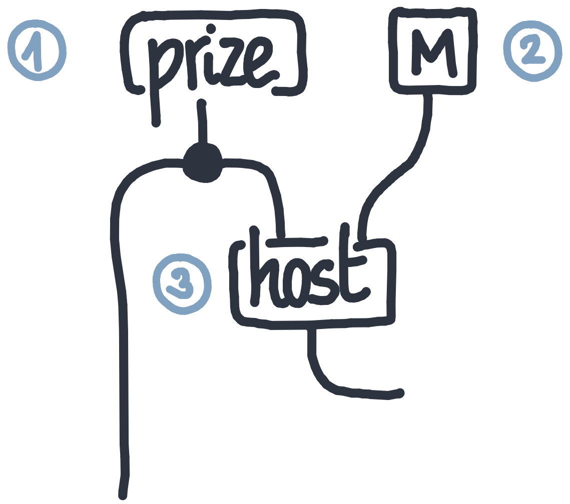


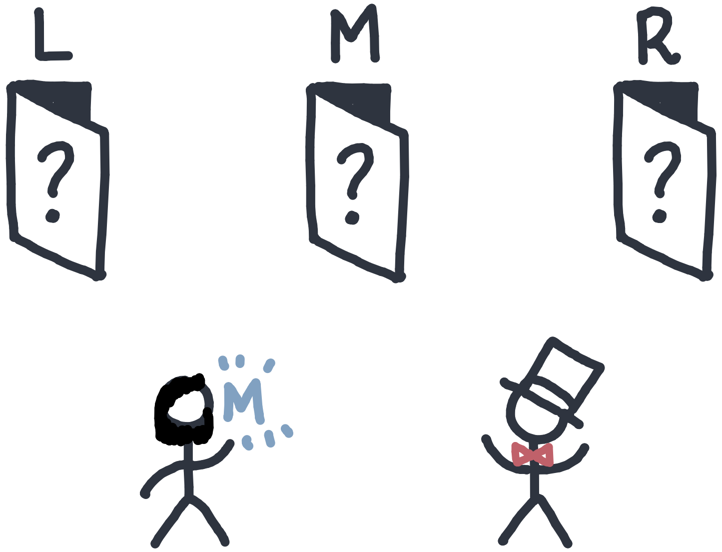


① $\frac{1}{3}|L\rangle + \frac{1}{3}|M\rangle + \frac{1}{3}|R\rangle$

② $\frac{1}{3}|L,M\rangle + \frac{1}{3}|M,M\rangle + \frac{1}{3}|R,M\rangle$

MONTY-HALL PROBLEM



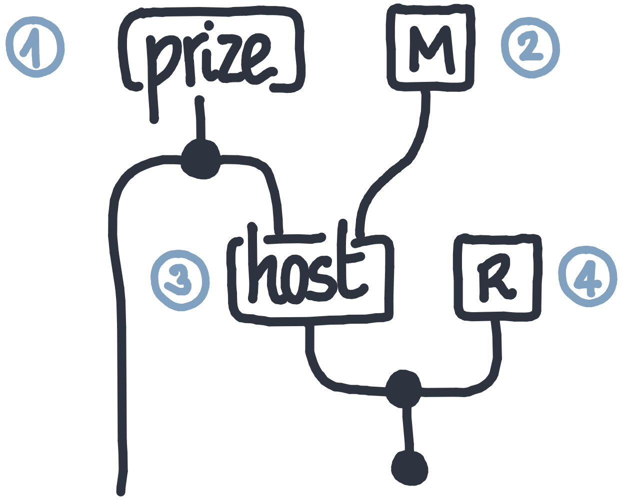


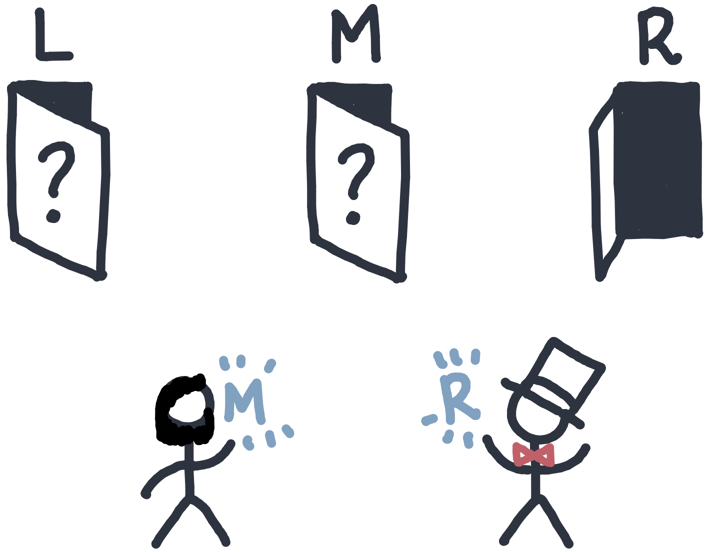
① $\frac{1}{3}|L\rangle + \frac{1}{3}|M\rangle + \frac{1}{3}|R\rangle$

② $\frac{1}{3}|L,M\rangle + \frac{1}{3}|M,M\rangle + \frac{1}{3}|R,M\rangle$

③ $\frac{1}{3}|L,M,R\rangle + \frac{1}{6}|M,M,R\rangle + \frac{1}{6}|M,M,L\rangle + \frac{1}{3}|R,M,L\rangle$

MONTY-HALL PROBLEM





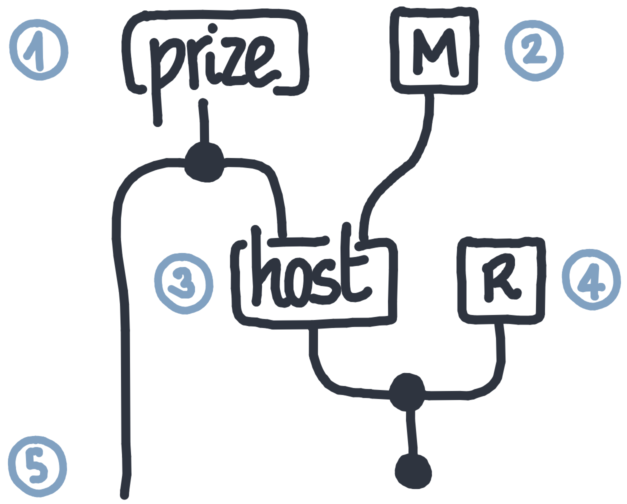
① $\frac{1}{3}|L\rangle + \frac{1}{3}|M\rangle + \frac{1}{3}|R\rangle$

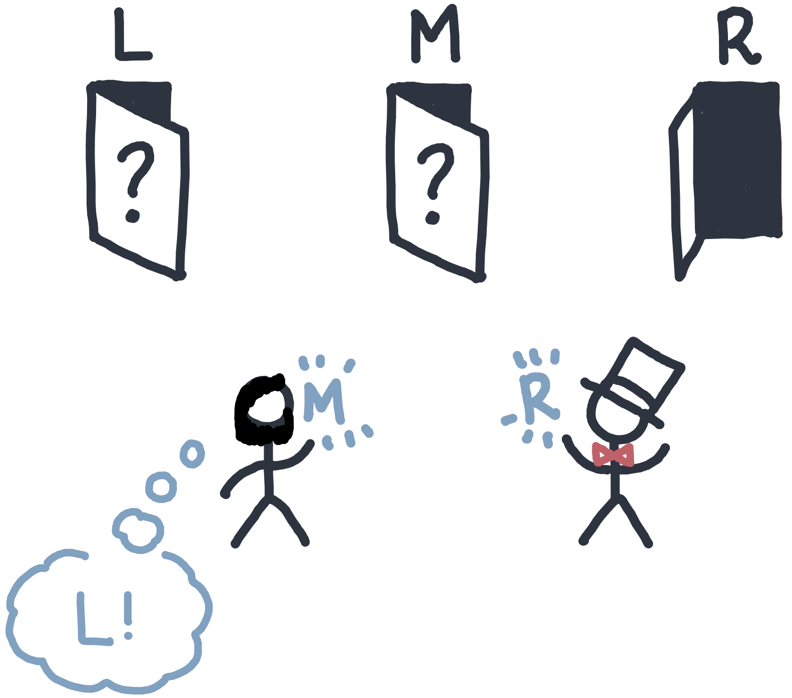
② $\frac{1}{3}|L,M\rangle + \frac{1}{3}|M,M\rangle + \frac{1}{3}|R,M\rangle$

③ $\frac{1}{3}|L,M,R\rangle + \frac{1}{6}|M,M,R\rangle$
$+ \cancel{\frac{1}{6}|M,M,L\rangle} + \cancel{\frac{1}{3}|R,M,L\rangle}$

④ $\frac{1}{3}|L,M,R\rangle + \frac{1}{6}|M,M,R\rangle$

MONTY-HALL PROBLEM





① $\frac{1}{3}|L\rangle + \frac{1}{3}|M\rangle + \frac{1}{3}|R\rangle$

② $\frac{1}{3}|L,M\rangle + \frac{1}{3}|M,M\rangle + \frac{1}{3}|R,M\rangle$

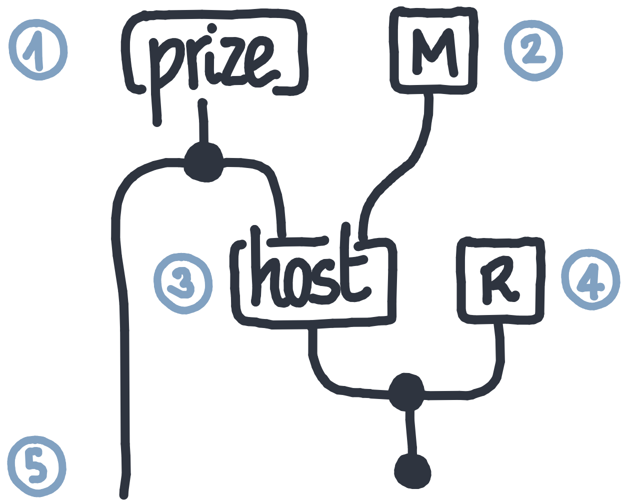
③ $\frac{1}{3}|L,M,R\rangle + \frac{1}{6}|M,M,R\rangle + \cancel{\frac{1}{6}|M,M,L\rangle} + \cancel{\frac{1}{3}|R,M,L\rangle}$

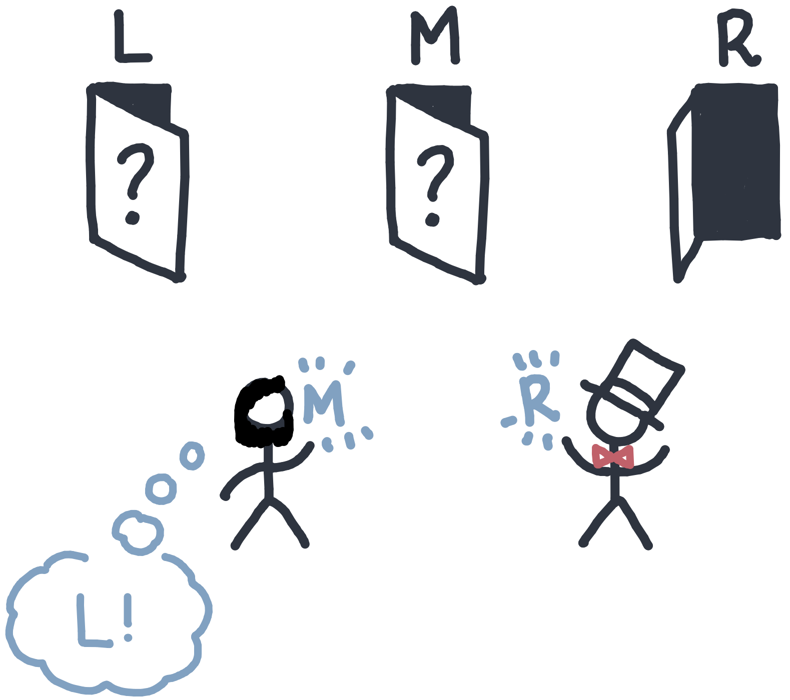
④ $\frac{1}{3}|L,M,R\rangle + \frac{1}{6}|M,M,R\rangle$

⑤ $\frac{1}{3}|L\rangle + \frac{1}{6}|M\rangle$

$\frac{2}{3}|L\rangle + \frac{1}{3}|M\rangle$

MONTY-HALL PROBLEM





```
montyHall :: Distribution Door
montyHall = do
① prize ← uniform [L,M,R]
② choice ← |M⟩
③ announcement ← host(prize, choice)
④ observe (announcement = R)
⑤ return (prize)
```

github.com/mroman42/observe

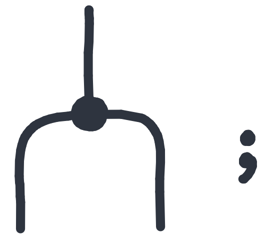
OUTLINE.

1. Copy-discard Categories.
2. Partial Markov Categories.
3. Discrete Partial Markov Categories.
4. Continuous Examples.

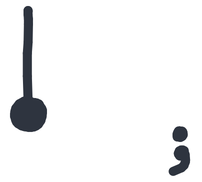
OUTLINE.

1. Copy-discard Categories.
2. Partial Markov Categories.
3. Discrete Partial Markov Categories.
4. Continuous Examples.

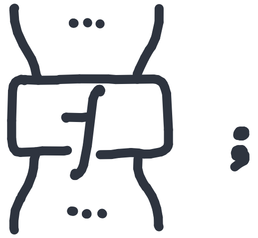
1. Copy-Discard Categories



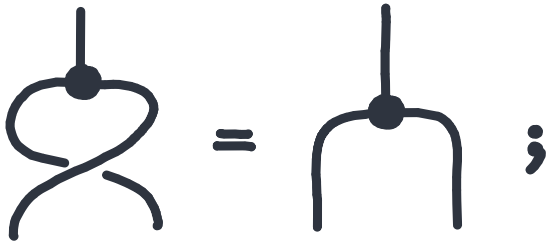
copy



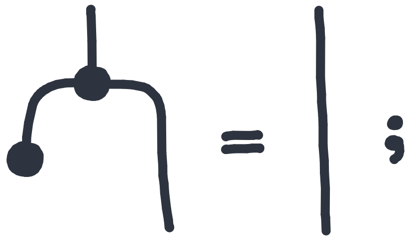
discard



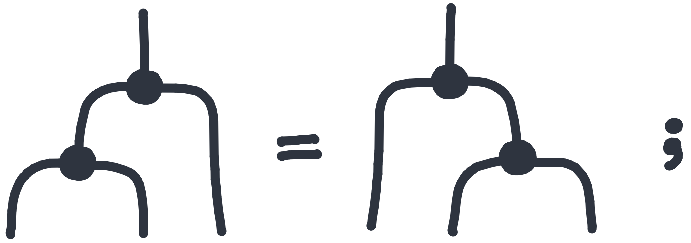
node



commutativity

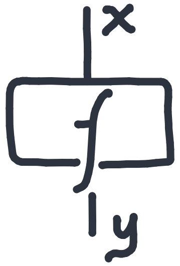


unitality



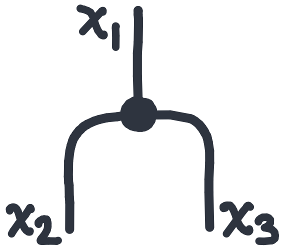
associativity

1. COPY-DISCARD CATEGORIES



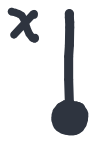
(sub) distribution

$\sum_y f(y|x) \leq 1$.



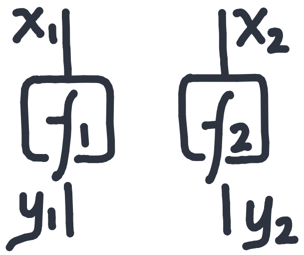
copy

$\delta(x_1 = x_2 = x_3)$



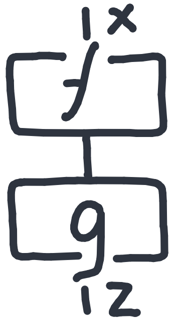
discard

1



monoidal tensor

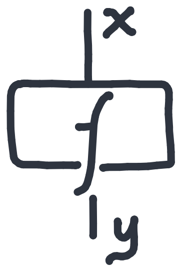
$f_1(y_1|x_1) \cdot f_2(y_2|x_2)$



composition

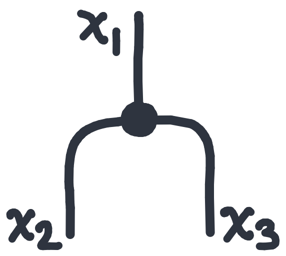
$\sum_y g(z|y) \cdot f(y|x)$.

1. Copy-Discard Categories



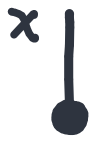
(sub) distribution

$\int_y f(dy|x) \leq 1$.



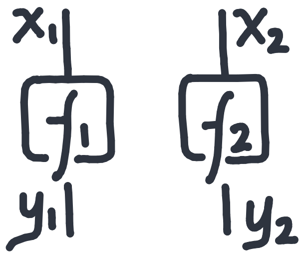
copy

$\delta_{x_1}(dx_2) \cdot \delta_{x_1}(dx_3)$



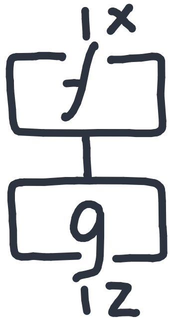
discard

1



monoidal tensor

$f_1(dy_1|x_1) \cdot f_2(dy_2|x_2)$



composition

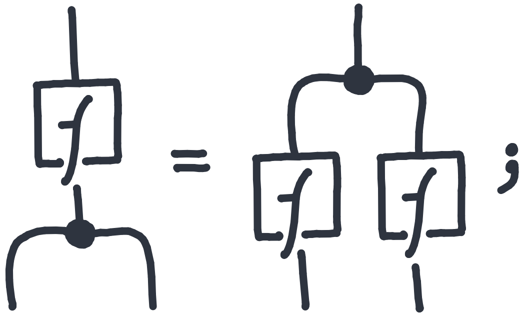
$\int_y g(dz|y) \cdot f(dy|x)$.

1. TOTAL & DETERMINISTIC

total

$$\sum_y f(y|x) = 1.$$

A (sub)distribution is total if it is a 'distribution' with measure 1.

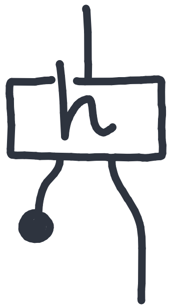


deterministic

$$f(y_1|x)\cdot f(y_2|x) = f(y|x)\cdot \delta(y=y_1=y_2).$$

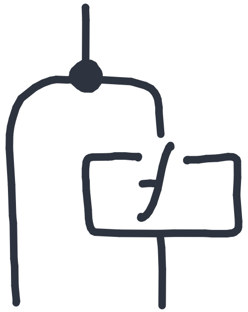
A (sub)distribution is deterministic when it is a partial function. $f(y|x)=1$ or $f(y|x)=0$.

1. MARGINALS & CONDITIONALS

$\pi_2(h) =$  ;

marginal

$\sum_y h(y,z|x)$

$\nu(f) =$  ;

input-copy

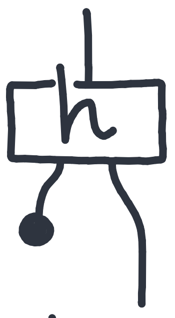
$f(y|x)\cdot\delta_{x'}(x).$

Section/retraction pair.

$\pi_2(\nu(f)) = f$

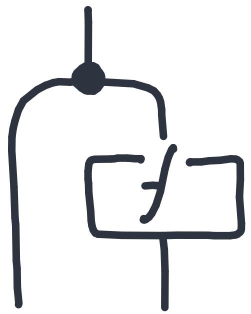
$\nu(\pi_2(h)) \neq h$

1. MARGINALS & CONDITIONALS

$\pi_2(h) =$


;

marginal

$\nu(f) =$


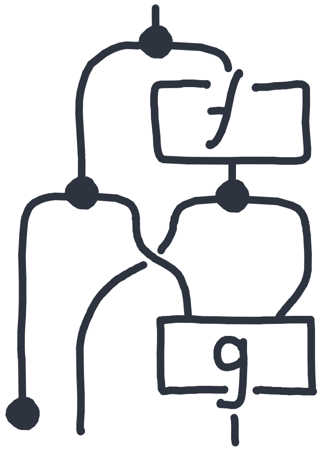
;

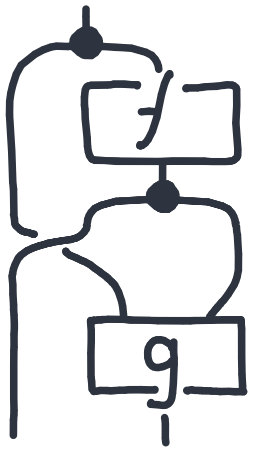
input-copy

Section/retraction pair.

$\pi_2(\nu(f)) = f$
$\nu(\pi_2(h)) \neq h$

DEFINITION. Conditional composition is composition conjugated by marginals.

$(f \blacktriangleleft g) =$


$=$


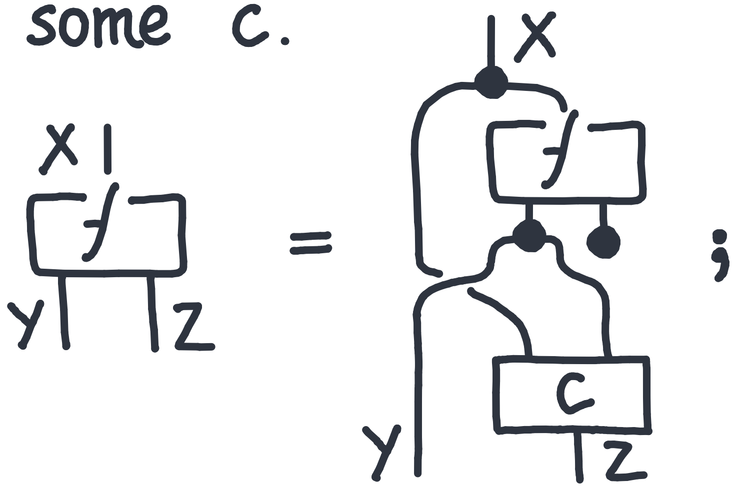
;

conditional composition

$f \blacktriangleleft g = \pi_2(\nu(f) ; \nu(g))$
$\varepsilon = \pi_2(id)$.

1. MARGINALS & CONDITIONALS

A morphism has conditionals if it can be conditionally-factored through its marginals: $f = (\pi_2 f) \blacktriangleleft c$ for some c.



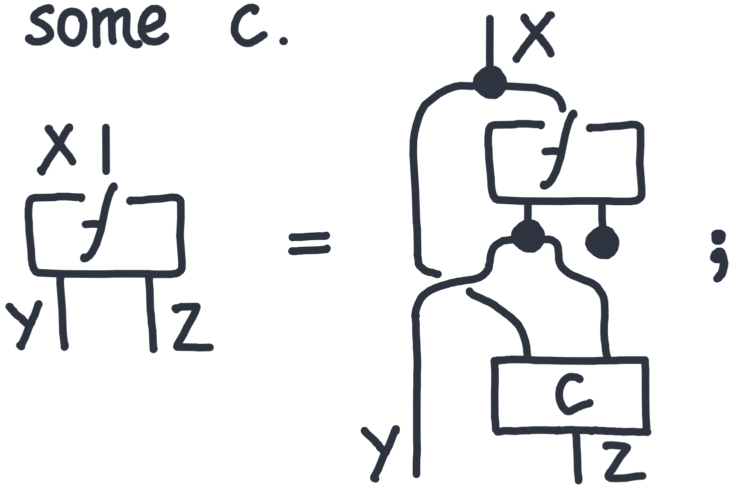
$$f(y,z|x) = \Sigma_{z_0} f(y,z_0|x) \cdot c(z|x,y);$$

$$c(z|x,y) = \frac{f(y,z|x)}{\Sigma_{z_0} f(y,z_0|x)}.$$

unless zero

1. ALMOST-SURE EQUALITY

A morphism has conditionals if it can be conditionally-factored through its marginals: $f = (\pi_2 f) \triangleleft c$ for some c.



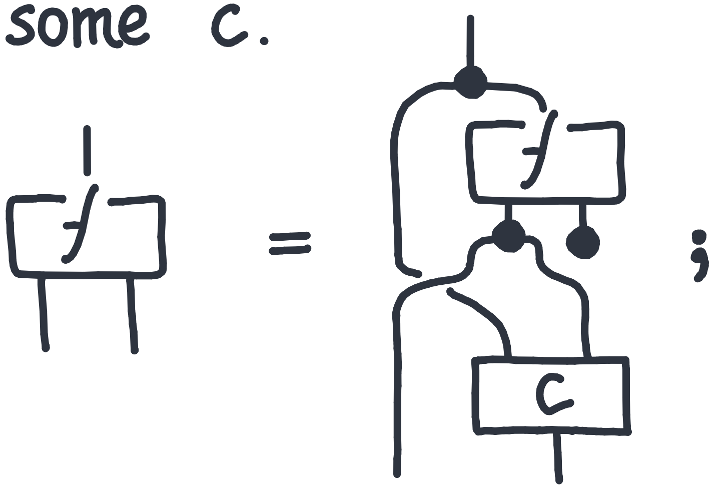
DEFINITION. Two morphisms $c, d : X \otimes Y \to Z$ are f-almost surely equal when $f \triangleleft c = f \triangleleft d$.

PROPOSITION. Conditionals of f are f-almost surely unique.

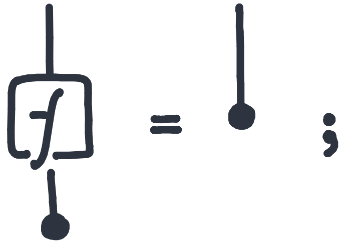
c.f. Fritz, 2020.

1. MARKOV CATEGORIES

A morphism has conditionals if it can be conditionally-factored through its marginals: $f = (\pi_2 f) \triangleleft c$ for some c.



DEFINITION. A Markov category is a copy-discard category where every morphism is total and has conditionals.



c.f. Fritz, 2020.

OUTLINE.

1. Copy-discard Categories.
2. Partial Markov Categories.
3. Discrete Partial Markov Categories.
4. Continuous Examples.

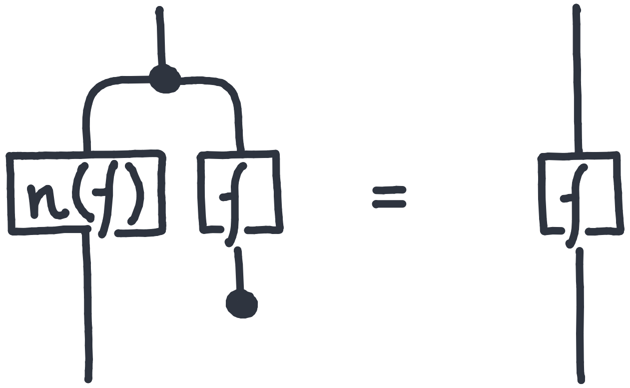
2. PARTIAL MARKOV CATEGORIES

DEFINITION. A partial Markov category is a copy-discard category where every morphism has conditionals.

- Domains.
- Normalization.
- Bayesian Inversion.

2. NORMALIZATION.

DEFINITION. A normalization of $f : A \to B$ is a morphism $n(f) : A \to B$ such that

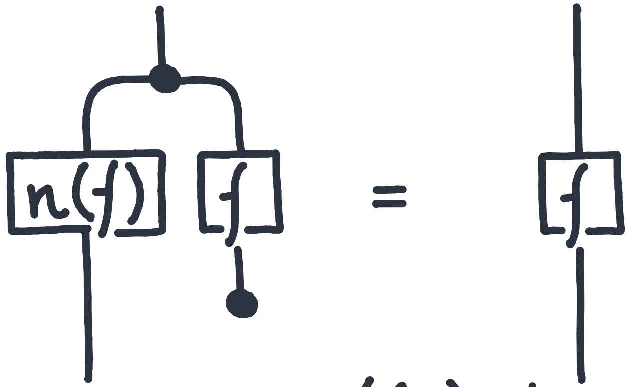


$$n(f)(y|x) \cdot \Sigma_{y'} f(y'|x) = f(y|x).$$

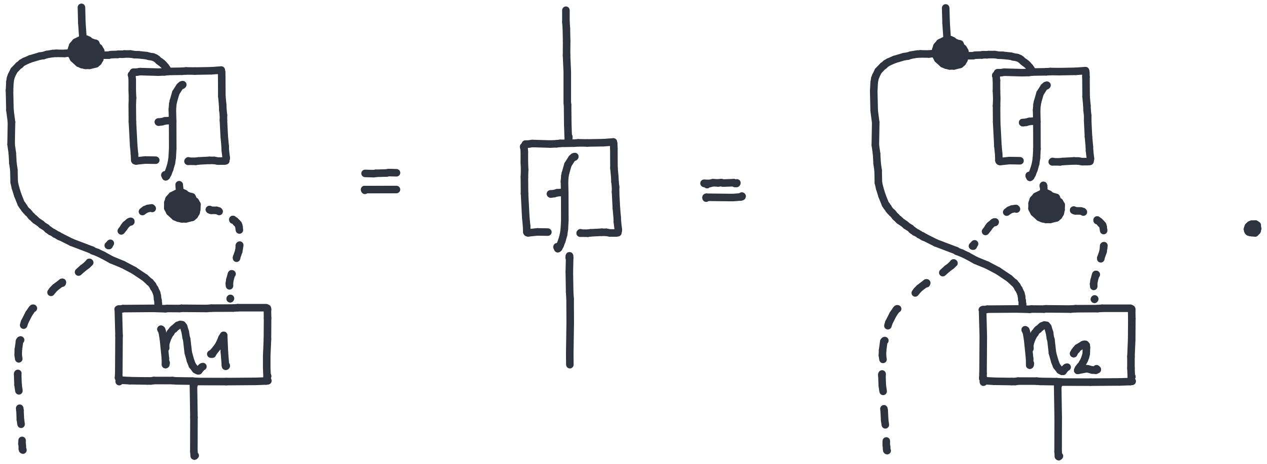
$$n(f)(y|x) = \frac{f(y|x)}{\Sigma_{y'} f(y'|x)}.$$

2. NORMALIZATION.

DEFINITION. A normalization of $f : A \to B$ is a morphism $n(f) : A \to B$ such that

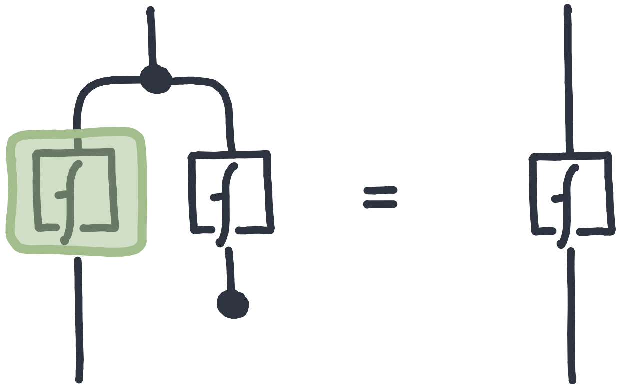


Normalization is not unique: it is $(f_{\bullet})$-almost surely unique.



2. NORMALIZATION.

DEFINITION. A normalization of $f: A \to B$ is a morphism $n(f): A \to B$ such that



$$n(f)(y|x) \cdot \Sigma_{y'} f(y'|x) = f(y|x).$$

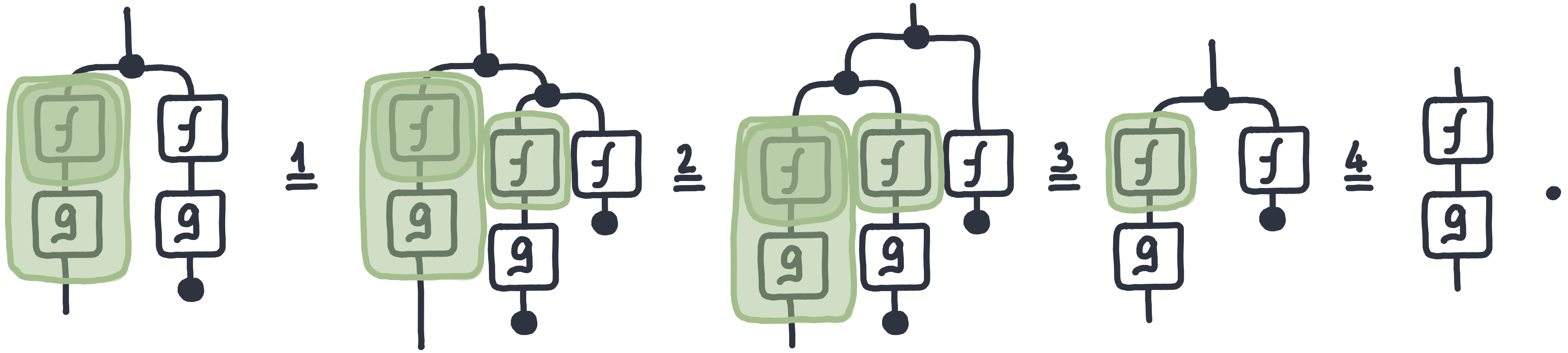
$$n(f)(y|x) = \frac{f(y|x)}{\Sigma_{y'} f(y'|x)}.$$

2. NORMALIZATION.

PROPOSITION. We can renormalize during a computation.

$$n(n(f) \fatsemi g) \approx_{(f g \bullet)} n(f \fatsemi g).$$

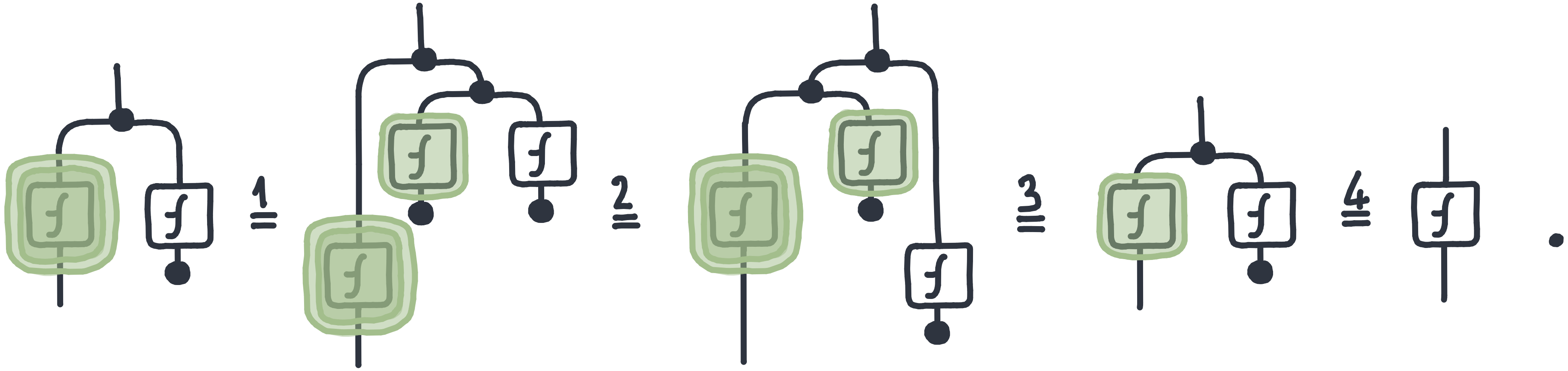
PROOF. By definition of normalization.



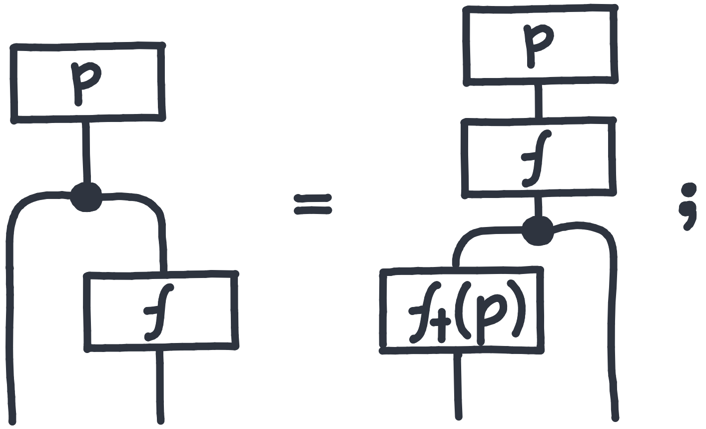
2. NORMALIZATION.

PROPOSITION. Normalization is $(f_\bullet)$-almost surely idempotent,

$$n(n(f)) \approx_{(f_\bullet)} n(f).$$



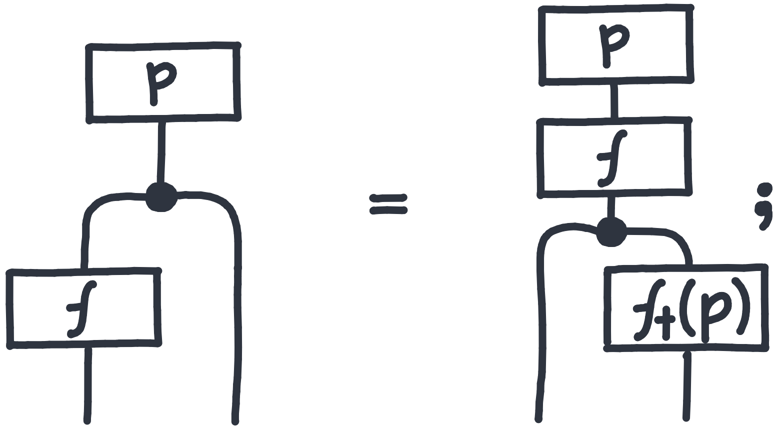
2. BAYESIAN INVERSION.



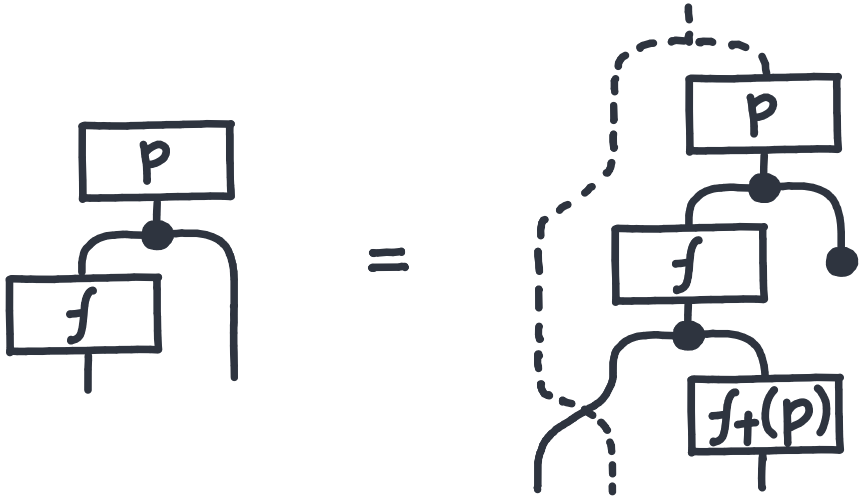
$$p(x)\cdot f(y|x) = \sum_{x'} p(x')\cdot f(y|x')\cdot f_{\dagger}(p)(x|y).$$

$$f_{\dagger}(p)(x|y) = \frac{p(x)\cdot f(y|x)}{\sum_{x'} p(x')\cdot f(y|x')}$$

2. BAYESIAN INVERSION.



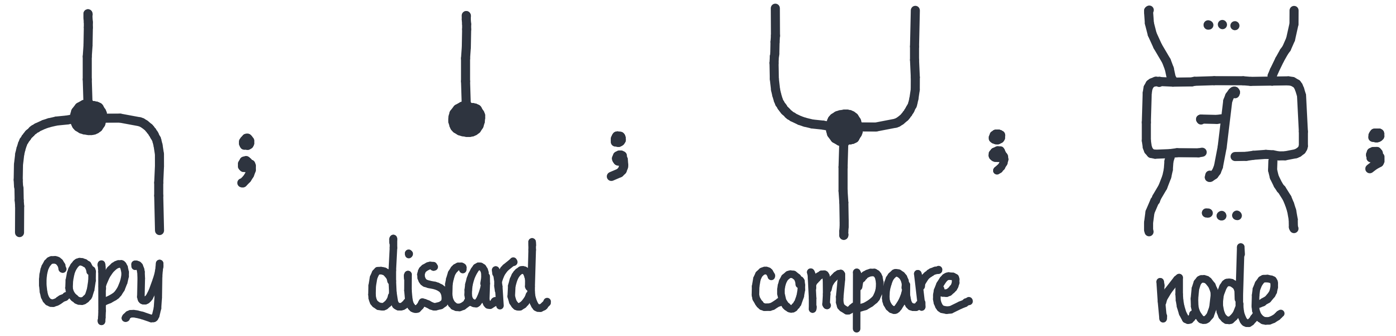
LEMMA. Bayesian inversions exist in any partial Markov category and are (pf)-almost surely unique: they are particular cases of conditionals.

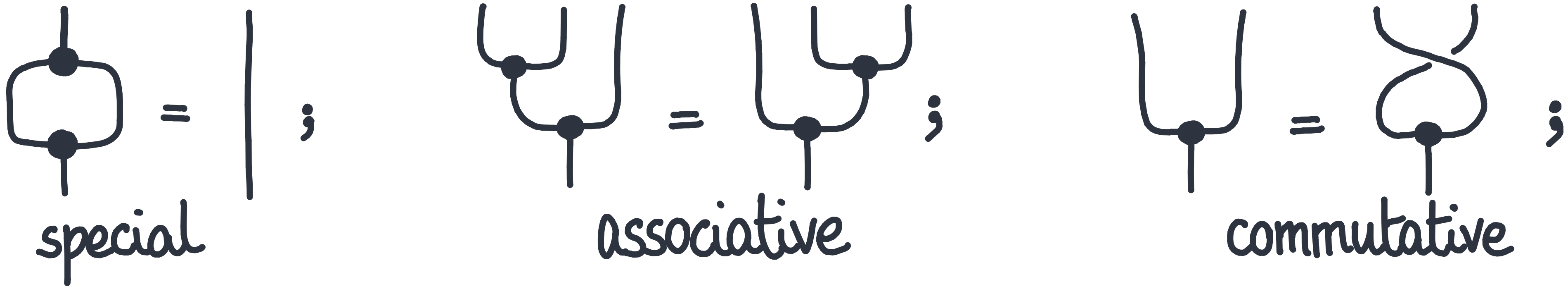


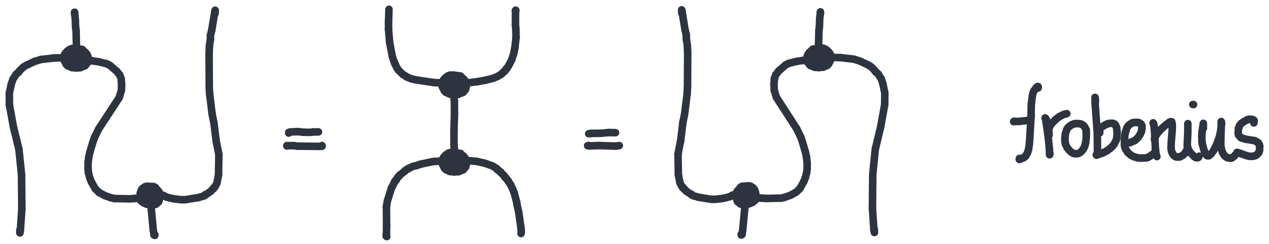
OUTLINE.

1. Copy-discard Categories.
2. Partial Markov Categories.
3. Discrete Partial Markov Categories.
4. Continuous Examples.

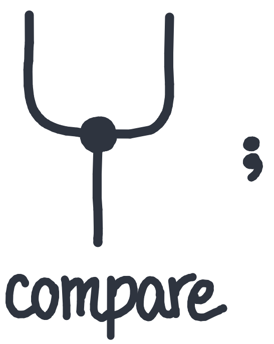
3. DISCRETE PARTIAL MARKOV CATEGORIES





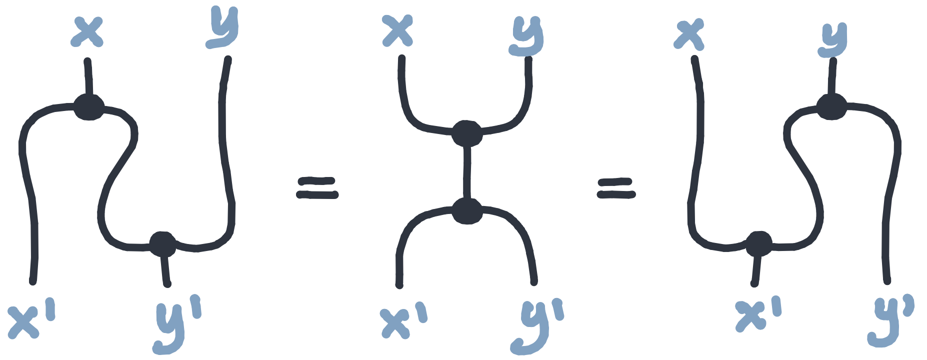


3. DISCRETE PARTIAL MARKOV CATEGORIES



$m(x|x_1,x_2) = [x_1=x_2=x]$

not a total function.

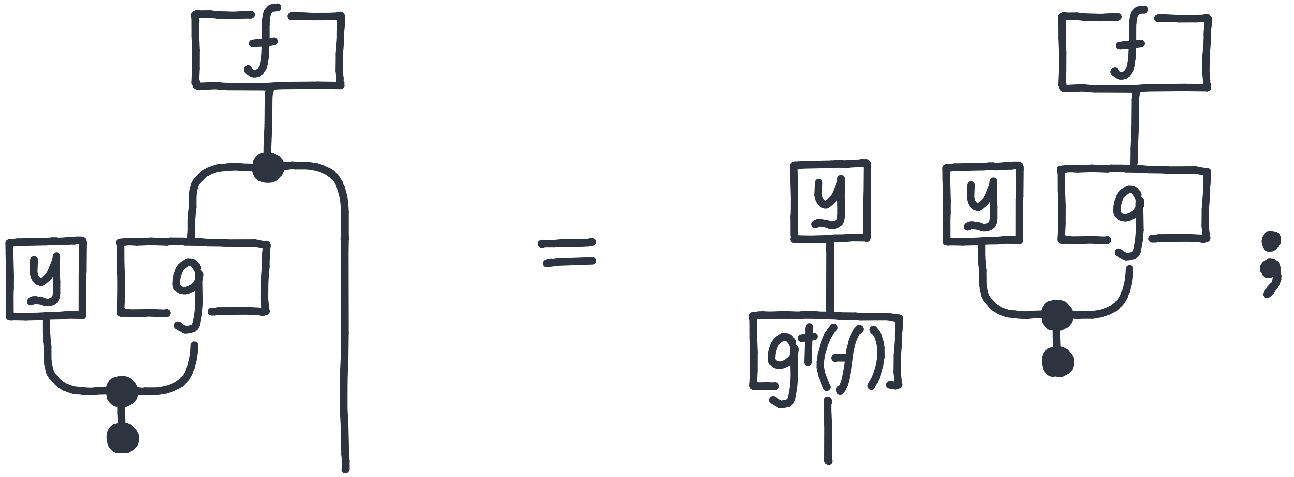


$$\delta(x=y=y')\cdot\delta(x=x')$$
$$=$$
$$\textstyle\sum_z \delta(x=y=z)\cdot\delta(z=x'=y').$$

3. BAYES' THEOREM

$P(x|y) = P(y|x)\cdot P(x)/P(y).$

Observing $y \in Y$ through a channel $f : X \to Y$ from a prior $p : 1 \to X$ updates the prior, up to scalar, to the Bayesian inversion evaluated on the observation.



$$f(x)\cdot g(y|x) = g^{\dagger}(f)(x|y)\cdot \Sigma_x f(x)\cdot g(y|x) ;$$

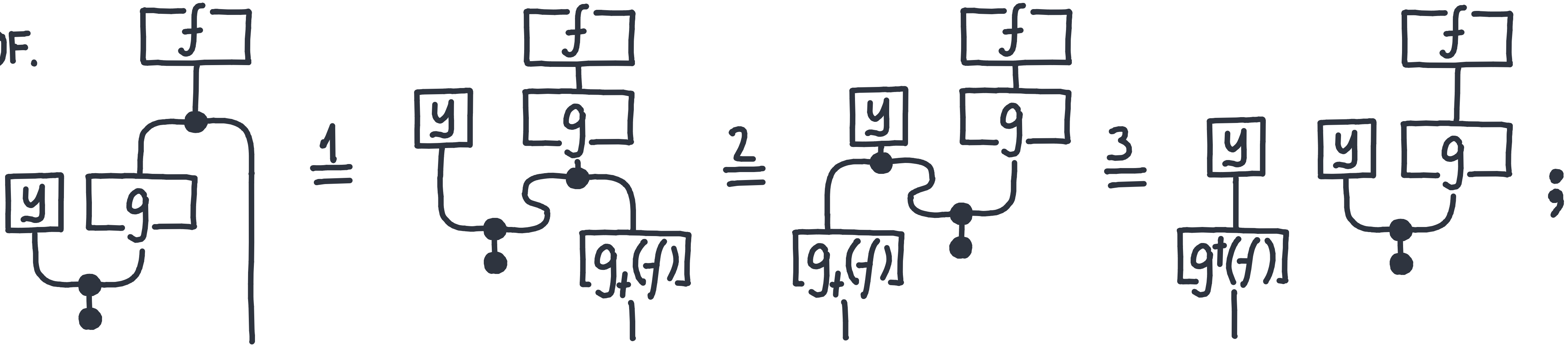
$$g^{\dagger}(f)(x|y) = \frac{f(x)\cdot g(y|x)}{\Sigma_x f(x)\cdot g(y|x)} ;$$

→ unless zero.

3. BAYES' THEOREM

Observing $y \in Y$ through a channel $f : X \to Y$ from a prior $p : 1 \to X$ updates the prior, up to scalar, to the Bayesian inversion evaluated on the observation.

PROOF.



OUTLINE.

1. Copy-discard Categories.
2. Partial Markov Categories.
3. Discrete Partial Markov Categories.
4. Continuous Examples.

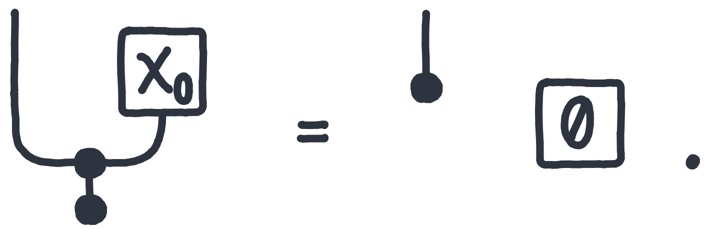
4. CONTINUOUS PARTIAL MARKOV CATEGORIES

Continuous probability categories have comparators, but not useful.

$$\varphi(A|x,y) = \begin{cases} 1 & \text{when } x=y\in A, \\ 0 & \text{otherwise.} \end{cases}$$

is a measurable function for Standard Borel spaces.

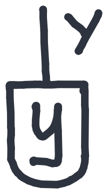
However, the set $\{x_0\}$ has measure zero: it is impossible to get exactly x_0.



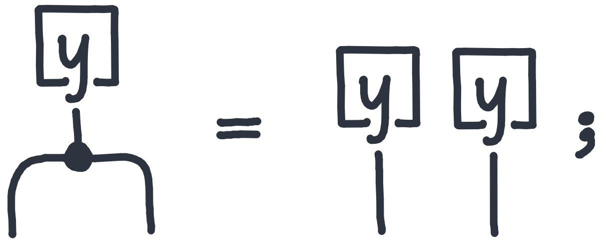
This is bad: is there a way to keep exact observations?

4. CONTINUOUS PARTIAL MARKOV CATEGORIES

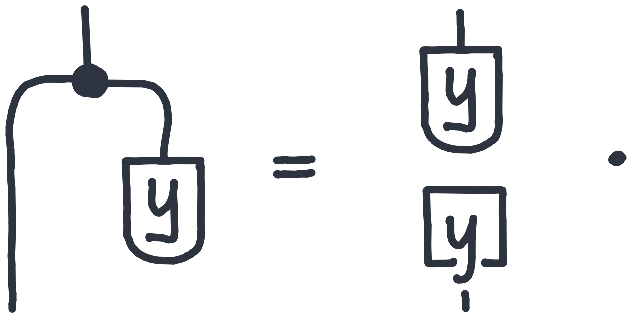
DEFINITION. Given a Markov category $\mathbb{C}$, we construct a partial Markov category, $\mathrm{exact}(\mathbb{C})$, that adds a morphism $y^\circ : Y \to I$ for every deterministic $y : I \to Y$,



for each



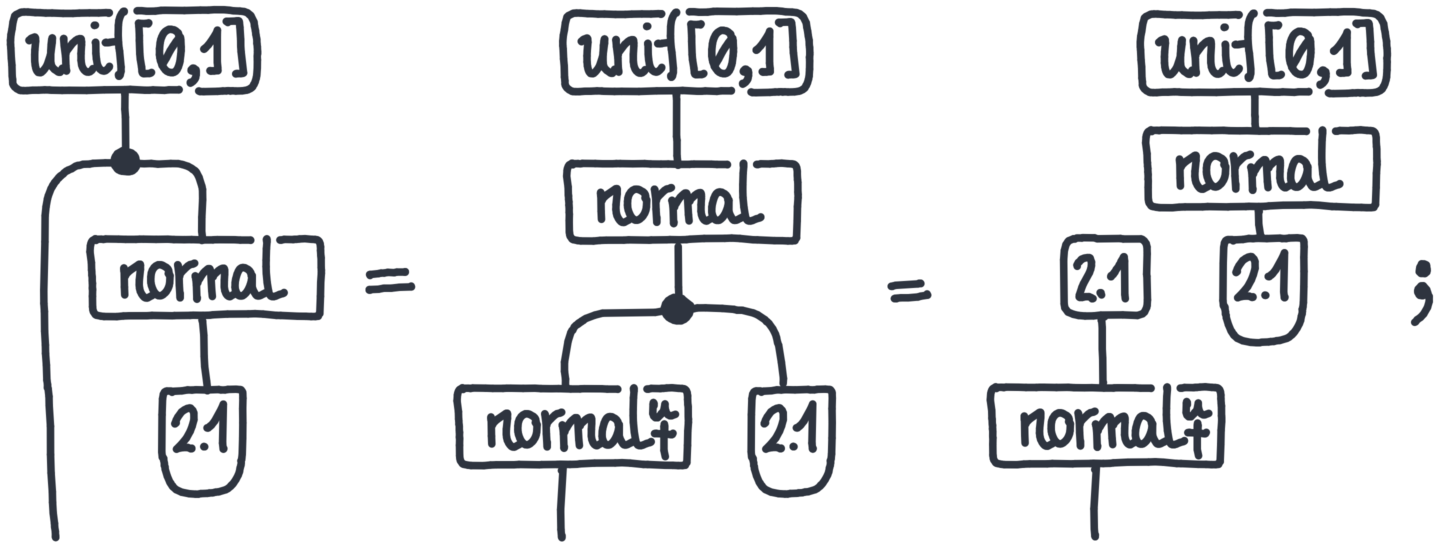
quotiented by the equation

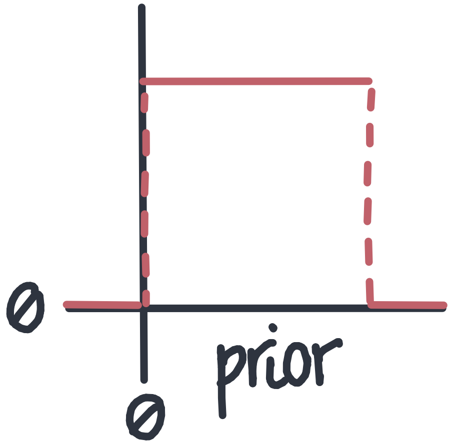


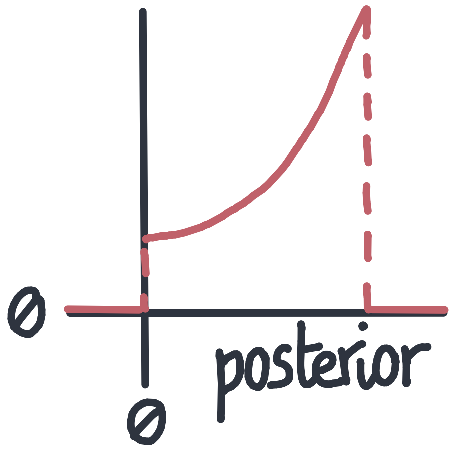
c.f. Staton, Stein.

4. CONTINUOUS PARTIAL MARKOV CATEGORIES

EXAMPLE. Assume a normal distribution with s.d. 1 and mean sampled uniformly from [0,1]. Say we sample a 2.1 out of this normal, what is the mean?



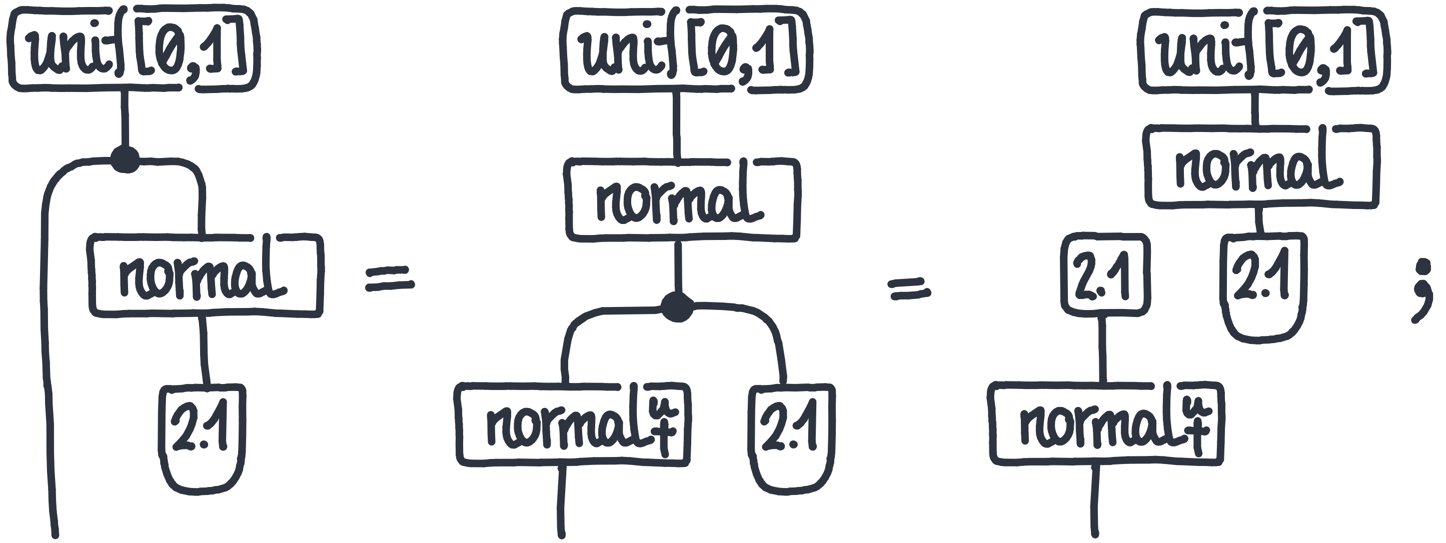


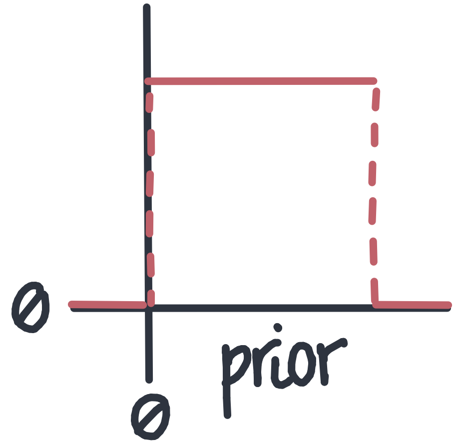


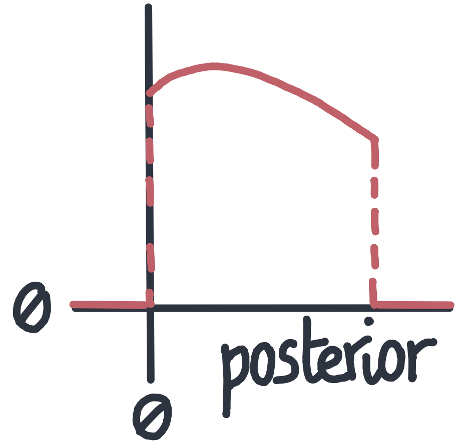
$$\text{normal}_{\dagger}(\text{unif})(m|2.1) = \frac{\text{unif}_{0,1}(m) \cdot \text{normal}(2.1|m)}{\int_{m'} \text{unif}_{0,1}(m') \cdot \text{normal}(2.1|m')}$$

4. CONTINUOUS PARTIAL MARKOV CATEGORIES

EXAMPLE. Assume a normal distribution with s.d. 1 and mean sampled uniformly from [0,1]. Say we sample a 0.3 out of this normal, what is the mean?







$$\mathrm{normal}_{\dagger}(\mathrm{unif})(m|0.3) = \frac{\mathrm{unif}_{0,1}(m)\cdot \mathrm{normal}(0.3|m)}{\int_{m'} \mathrm{unif}_{0,1}(m')\cdot \mathrm{normal}(0.3|m')}$$

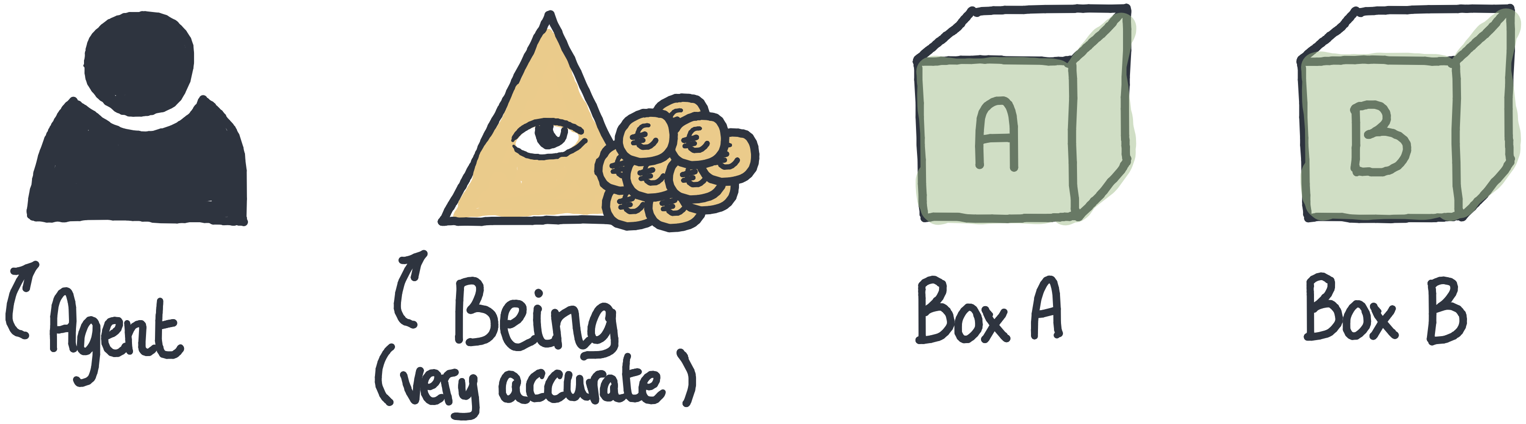
END

PART 0: DECIDING IS DIFFICULT

‹NEWCOMB'S PROBLEM›

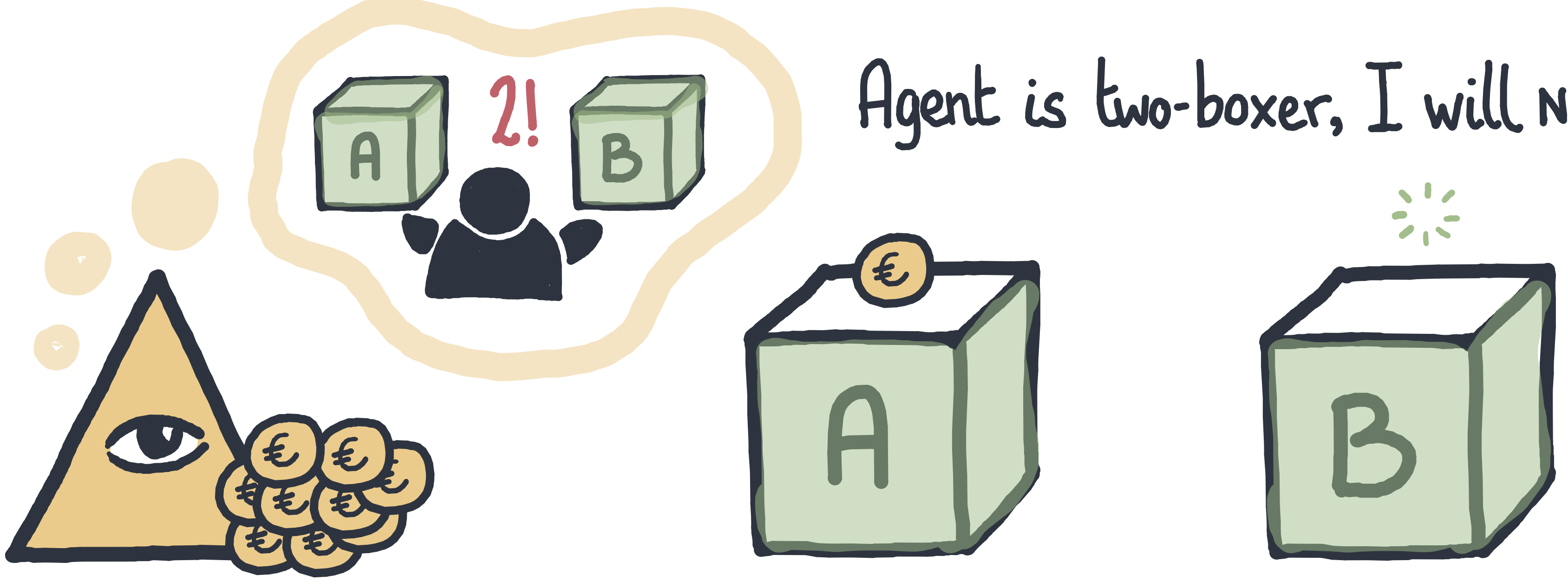
Nozick

NEWCOMB'S PROBLEM



NEWCOMB'S PROBLEM

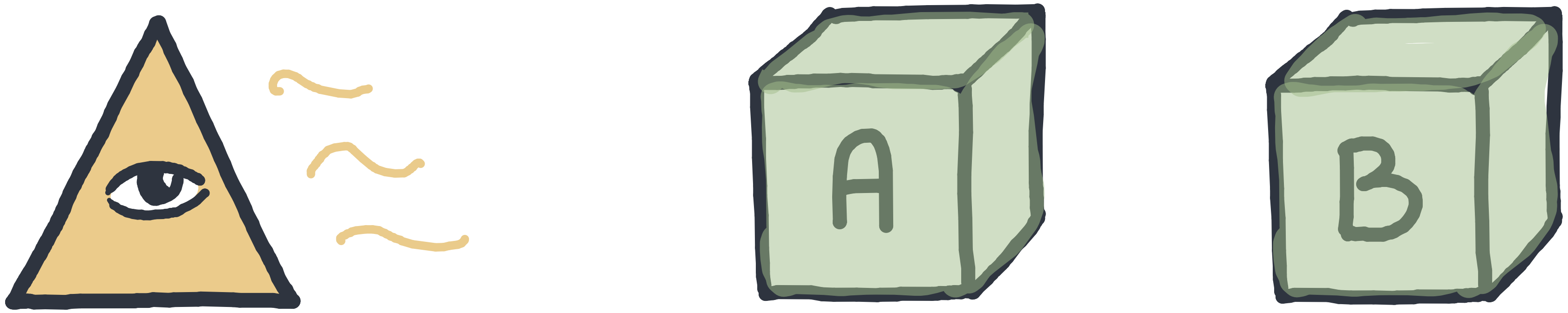
Agent is one-boxer, I will fill both boxes.



Agent is two-boxer, I will NOT fill both boxes.

NEWCOMB'S PROBLEM

Predictor closes the boxes & leaves.

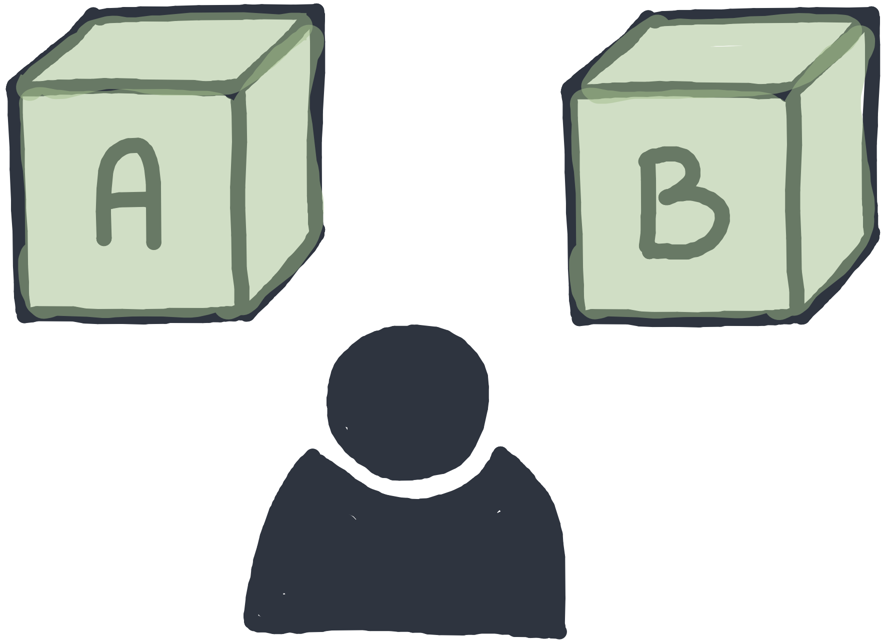


NEWCOMB'S PROBLEM

What should the agent do?

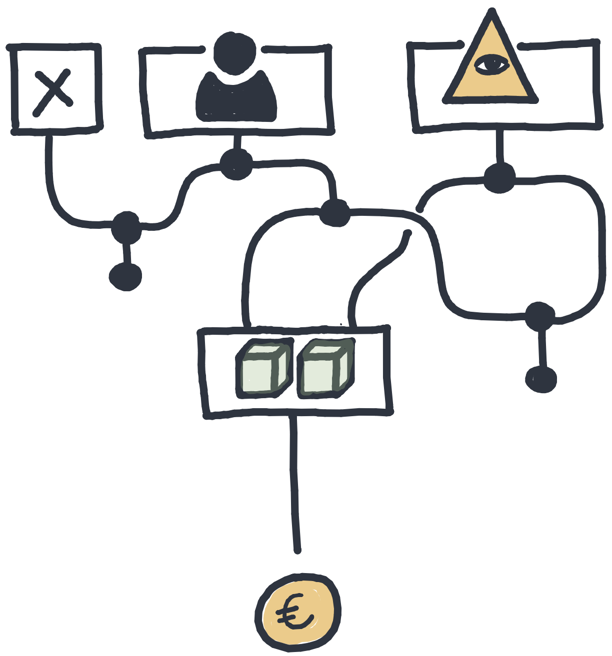
- Two-box: 1€ is better than nothing, 101€ is better than 100€. No matter what I do, I cannot change whatever is in the boxes.
- One-box: I want to one-box so that the accuracy of the predictor means that I get 100€, instead of 1.

The idea is always to find the argument that maximizes some expected value function. The debate is in interpreting what that function is.



CALCULEMUS

Leibniz's dream was to see philosophical disputes reduced to mathematical calculation. An algorithm for figuring out the correct position given assumptions.



Wishlist.

- Formal syntax and axioms for stochastic processes and bayesian inference.
- Systematic decision theory.
- Compositional and abstract theory extending synthetic probability.

CALCULEMUS

Leibniz's dream was to see philosophical disputes reduced to mathematical calculation. An algorithm for figuring out the correct position.

```
do
  prediction ← △
  action ← 👤
  observe (action = x)
  observe (action = prediction)
  return (▢▢)(action, prediction)
```

Wishlist.

- Formal syntax and axioms for stochastic processes and bayesian inference.
- Systematic decision theory.
- Compositional and abstract theory extending synthetic probability.

End

REFERENCES

- **Fong.** A Categorical Perspective on Bayesian Networks.
- **Fritz.** A Synthetic Approach to Markov kernels.
- **Cho, Jacobs.** Disintegration and Bayesian inversion.
- **Nozick.** Newcomb's Problem and Two Principles of Choice.
- **Hughes.** Generalising Monads to Arrow.

SUMMARY

- Minimal algebra for evidential decision theory.
- Intuitive diagramatic syntax.
- Translating to actual code.
- Synthetically proving a Bayes' theorem.

Partial Markov categories extend synthetic probability algebra to allow observations.