Accepting Normalization via Markov Magmoids

Elena Di Lavore
Mario Román

Abstract

Normalization is not a distributive. We introduce distributive sesquilaws to abstract normalization and, from their axioms, we derive some of its properties. In particular, normalized kernels form a unital magmoid with an action of the category of subdistributions; its possibilistic analogue is the action from may-must relations into the category of relations. We argue that the magmoid of normalized kernels is the stochastic analog of the category of relations.

Keywords: Categorical semantics, Markov categories, probabilistic programming, tricocycloids.

1 Introduction

1.1 The Quest for Normalized Kernel Composition

Imagine solving the Monty-Hall problem [vS]. We *(i)* consider the prior probability that a prize is behind any of three doors; then, we know that *(ii)* after choosing the middle door, the host will open a door that cannot be neither the chosen door nor the one with the prize; *(iii)* we will observe that the host opens the middle door; and finally, *(iv)* we will renormalize the probabilities.

$$\begin{aligned}
&\overset{(i)}{\rightsquigarrow} && 1/3\,|L\rangle + 1/3\,|M\rangle + 1/3\,|R\rangle \\
&\overset{(ii)}{\rightsquigarrow} && 1/3\,|LR\rangle + 1/6\,|ML\rangle + 1/6\,|MR\rangle + 1/3\,|RL\rangle \\
&\overset{(iii)}{\rightsquigarrow} && 1/3\cancel{|LR\rangle} + 1/6\,|ML\rangle + 1/6\cancel{|MR\rangle} + 1/3\,|RL\rangle \\
&\overset{(iv)}{\rightsquigarrow} && 1/3\,|ML\rangle + 2/3\,|RL\rangle\,.
\end{aligned}$$

The phenomenon we seek to study occurs at the last two steps. We first *(iii)* obtain something that is not a full distribution, but only a subdistribution; and *(iv)* we multiply everything by a constant — by 2, in this example — to obtain again a distribution. This interplay between distributions and subdistributions usually leads us to pick substochastic kernels for probabilistic programming semantics: functions $X \to DMY$, for D the distribution monad and M the maybe monad. These include both distributions and subdistributions.

However, we may wish to avoid subdistributions: if the last two steps were compressed into one, there would be no need for subdistributions. Instead, we could work with normalized kernels, $X \to MDY$, consisting of either nothing or a full distribution. May we compose normalized kernels? The quest for a category of normalized kernels has brought us multiple techniques; let us review some here.

(a) Work up-to-scalar. Two subdistributions, $d_1, d_2 \in DMX$, share their normalization whenever there exists some positive real number, $\lambda \in \mathbb{R}^+$, such that $d_1(x) = \lambda \cdot d_2(x)$. Working up-to-scalar means bringing this idea to kernels: we identify two kernels $f_1, f_2 \colon X \to DMX$ up to scalar multiplication [SS24, Definition 6.3; PTRSZ25, Definition 2.9]; that is, whenever there exists some positive real, $\lambda \in \mathbb{R}^+$, such that

$$f_1(x; y) = \lambda \cdot f_2(x; y).$$

However, unlike subdistributions, substochastic kernels do have different normalization constants depending on their input: this quotient does not identify a kernel with its normalization.

(b) Work up-to-parameterized-scalar. The obvious solution is to vary the normalization constant based on the input to the stochastic kernel. In practice, this means identifying two kernels $f_1, f_2 \colon X \to DMX$ whenever there exists some family of positive reals, $\lambda(x) \in \mathbb{R}^+$, satisfying

$$f_1(x; y) = \lambda(x) \cdot f_2(x; y).$$

This quotienting does bring us to normalized channels, $X \to MDY$, but it stops being preserved by composition. In the abstract setting, the normalized morphisms of any partial Markov category are not closed under composition [DR23, Definitions 3.1 and 3.19].

(c) Work up-to-failure. A more radical solution does yield a category: the so-called *black-hole semantics* [Fri09, SW18]. Black-hole semantics arises from a valid distributive law casting any subdistribution into a distribution,

$$\mathrm{cast} \colon DMX \to MDX.$$

The problem is that this distributive law is defined by returns the failure element whenever we do not have a full distribution.

$$\mathrm{cast}(d) = \begin{cases} \bot, & \text{if } d(\bot) > 0; \\ d, & \text{otherwise.} \end{cases}$$

Alas, this distributive law is not helpful here: we would lose the solution to any probabilistic problem that involves any subdistribution at any point, for it would automatically be equated to failure.

The Kleisli category of this distributive law is the category of partial stochastic kernels, $X \to MDY$, which yields failure whenever possible. In the abstract setting, partial stochastic kernels have been developed as the leading example of quasi-Markov categories [FGL⁺25, Sha25].

Accepting Normalization. None of these solutions really constructs a category of normalized kernels. Should we continue this quest? Have we missed further solutions? Fortunately, Sokolova and Woracek classified all possible single-point extensions of distributions [SW18, Theorem 5.1] into (*i*) those with *imitating behaviour*, which identify failure and an existing element but do not address normalization; (*ii*) those with *extremal point behaviour*, which, like subdistributions, introduce a failure element that does not interact with the rest; (*iii*) those with *black-hole behaviour*, which, like partial distributions, make all elements adhere to failure.

Instead of trying to escape this classification, this paper argues that we may accept that normalized kernels do not form a category: the composition we need is indeed non-associative. At the same time, that a category structure may be unnecessary, after all, and that the behaviour of normalization may be better approximated by an action of the genuine category of subdistributions into a magmoid.

2 Normalization

Normalization is difficult to classify categorically. While it induces a natural transformation that braids the distribution (D) and maybe (M) monads, $\mathsf{n}_X \colon DMX \to MDX$, it is not a distributive law. While it induces an idempotent operation on substochastic channels, $\mathsf{n} \colon \mathsf{Subd}(X;Y) \to \mathsf{Subd}(X;Y)$, it is not functorial. While it induces a composition of normalized stochastic channels, $\mathsf{n} \colon \mathsf{Norm}(X;Y) \times \mathsf{Norm}(Y;Z) \to \mathsf{Norm}(X;Z)$, it is not associative.

Normalization of subdistributions into distributions is a fundamental operation of probability theory, but it is generally regarded as ill-behaved [Jac17].

Instead, we could argue its structure is rich: normalization induces a monoidal magmoid with copy-discard maps and conditionals; an almost distributive law interacting with the actual distributive law of subdistributions; and an action of the category of substochastic channels into normalized channels.

This paper takes a synthetic approach to normalization. We organize the algebra of normalization into multiple monoidal category-like structures — a Markov category, a partial Markov category, a quasi-Markov category, and a Markov magmoid — and derive all of it from a specialized interaction between a distributive law and an almost-distributive law: a distributive sesquilaw.

2.1 Normalization

Definition 1 (Normalization). *Normalization*, $\mathsf{n}_X \colon DMX \to MDX$, is a natural transformation defined by the following partial function

$$\mathsf{n}(f)(x) = \frac{f(x)}{\sum_{x' \in X} f(x')},$$

which is undefined, $\mathsf{n}(f) = \bot$, whenever $\sum_{x' \in X} f(x') = 0$.

Normalization is monoidal. Indeed, both the finitary distribution monad and the maybe monad are monoidal monads: their Kleisli categories, Stoch and Par, are both copy-discard categories. Normalization inherits this compatibility.

Proposition 2 (Normalization is monoidal). *Normalization of two distributions is the normalization of their joint independent distribution,* $\mathsf{n}(f \otimes g) = \mathsf{n}(f) \otimes \mathsf{n}(g)$.

$$\frac{f(x) \cdot g(y)}{\sum_{u \in X, v \in Y} f(u) \cdot g(v)} = \frac{f(x)}{\sum_{u \in X} f(u)} \cdot \frac{g(y)}{\sum_{v \in Y} g(v)}.$$

Proof. By calculation, or the discrete Fubini theorem.

$$\begin{aligned} \mathsf{n}(f \otimes g)(x,y) &= \frac{f(x) \cdot g(y)}{\sum_{u \in X, v \in Y} f(u) \cdot g(v)} \\ &= \frac{f(x)}{\sum_{u \in X} f(u)} \cdot \frac{g(y)}{\sum_{v \in Y} g(v)} \\ &= \mathsf{n}(f) \otimes \mathsf{n}(g). \end{aligned}$$ □

Were normalization to form a distributive law, its Kleisli category, Norm, would also be monoidal. The surprising fact is that normalization fails to form a distributive law, and this potential Kleisli category is instead a Kleisli magmoid.

2.2 The normalization magmoid

Definition 3 (Unital magmoid). A *unital magmoid* — or, non-associative category — consists of a collection of objects, $\mathbb{A}_{obj}$, and a set of morphisms, $\mathbb{A}(X;Y)$, for each two objects, $X, Y \in \mathbb{A}_{obj}$, endowed with — for each $X, Y, Z \in \mathbb{A}_{obj}$ — composition and identity operations

$$\begin{aligned} &(⨾) \colon \mathbb{A}(X;Y) \times \mathbb{A}(Y;Z) \to \mathbb{A}(X;Z), \text{ and} \\ &\mathrm{id} \colon \mathbb{A}(X;X); \end{aligned}$$

that are unital, meaning $f ⨾ \mathrm{id} = f = \mathrm{id} ⨾ f$.

Proposition 4 (Normalization magmoid). *Normalized stochastic channels between sets, $X \to MDY$, form a magmoid — the* normalized distribution magmoid, Norm *— where composition of two morphisms, $f \colon X \to MDY$ and $g \colon Y \to MDZ$, is defined as*

$$(f ⨾ g)(x;z) = \frac{\sum_{v \in Y} f(x;v) \cdot g(v;z)}{\sum_{v \in Y} \sum_{w \in Z} f(x;v) \cdot g(v;w)}.$$

In other words, if we consider the associated substochastic channels, $f^{\bullet} \colon X \to DMY$ and $g^{\bullet} \colon Y \to DMZ$, it is the normalization of their composition as subdistributions, $f ⨾ g = \mathsf{n}(f^{\bullet} \mathbin{;} g^{\bullet})$.

The two ways of associating this composition do give rise to different results. Arguably, left-associating composition behaves as expected,

$$((f ⨾ g) ⨾ h)(x;w) = \frac{\sum_{y,z} f(x;y) \cdot g(y;z) \cdot h(z;w)}{\sum_{y,z,w} f(x;y) \cdot g(y;z) \cdot h(z;w)}.$$

While right-associating composition may contain different normalization constants on the numerator and the denominator, making it impossible to simplify it.

$$(f \fatsemi (g \fatsemi h))(x;w) = \frac{\sum_y f(x;y) \cdot \frac{\sum_z g(y;z)\cdot h(z;w)}{\sum_{z,w} g(y;z)\cdot h(z;w)}}{\sum_{y,z} f(x;y) \cdot \frac{\sum_z g(y;z)\cdot h(z;w)}{\sum_{z,w} g(y;z)\cdot h(z;w)}}.$$

Proposition 5. *The normalized distribution magmoid is not a category.*

Still, even if not all morphisms associate, some of them do. All the useful structural morphisms, and all the deterministic partial functions of this magmoid do associate.

Definition 6 (Associating morphisms of a magmoid)**.** A morphism of a magmoid, $h \in \mathbb{A}(X;Y)$, is an *associating* morphism when

$$f \fatsemi (h \fatsemi g) = (f \fatsemi h) \fatsemi g$$

for any compatible pair of morphisms, $f \in \mathbb{A}(X';X)$ and $g \in \mathbb{A}(Y;Y')$.

Proposition 7 (Associating morphisms form a subcategory)**.** *Associating morphisms of a magmoid form a category with the composition of the original magmoid.*

The missing piece is to lift the monoidality of normalization to this newly constructed magmoid.

Definition 8 (Strict monoidal magmoid)**.** A *strict monoidal magmoid*, $\mathbb{A}$, consists of a monoid of objects, $(\mathbb{A}_{obj}, \otimes, I)$, and a collection of morphisms, $\mathbb{A}(X;Y)$, for each two objects, $X, Y \in \mathbb{A}_{obj}$. A strict monoidal magmoid is endowed with composition, identity, and tensoring operations,

$$(\otimes)\colon \mathbb{A}(X;Y) \times \mathbb{A}(X';Y') \to \mathbb{A}(X \otimes X'; Y \otimes Y');$$
$$(\fatsemi)\colon \mathbb{A}(X;Y) \times \mathbb{A}(Y;Z) \to \mathbb{A}(X;Z);$$

which must satisfy the following axioms.

1. $f \fatsemi \mathrm{id}_Y = f = \mathrm{id}_X \fatsemi f$;
2. $f \otimes \mathrm{id}_I = f = \mathrm{id}_I \otimes f$;
3. $f \otimes (g \otimes h) = (f \otimes g) \otimes h$;
4. $\mathrm{id}_X \otimes \mathrm{id}_Y = \mathrm{id}_{X \otimes Y}$;
5. $(f \fatsemi g) \otimes (f' \fatsemi g') = (f \otimes f') \fatsemi (g \otimes g')$.

Remark 9 (Coherence for monoidal magmoids). Monoidal magmoids are pseudomonoids of the 2-category of magmoids with magmoid functors and distributing natural transformations. By the coherence theorem for pseudomonoids [Ver17], every monoidal magmoid is equivalent to a strict one.

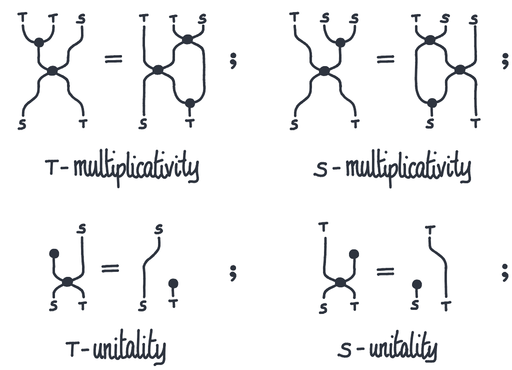
Proposition 10 (Norm is monoidal)**.** *The normalized distribution magmoid is monoidal with the cartesian product of sets and the following partial product of morphisms.*

$$(f_1 \otimes f_2)(x_1, x_2; y_1, y_2) = f_1(x_1;y_1) \cdot f_2(x_2;y_2).$$

3 Distributive Laws

Distributive laws [Bec69], their uses and limitations [ZM20], are all well-known. Briefly, the composition of two monads is not a monad again — in general, the tensor of two monoids is not a monoid again — but distributive laws endow this composition with monad structure.

Definition 11 (Distributive law [Bec69])**.** A *distributive law* between two monads, (S, μ, ν) and (T, μ, ν), on the same category is a natural transformation $\psi_X \colon TSX \to STX$ that moreover satisfies the following axioms.



Definition 12 (Monoidal distributive law)**.** A *monoidal distributive law* between two monoidal monads is a distributive law whose natural transformation is monoidal.

Proposition 13. *Given two monads, S and T, a distributive law between them induces a monad structure on the composite functor $S \circ T$. Given two monoidal monads, S and T, a monoidal distributive law between them induces a monoidal monad structure on the composite functor $S \circ T$.*

3.1 Subdistributions

Normalized channels can be composed inside a bigger category: the category of subdistributions, subStoch. There is indeed a monoidal distributive law, $MD \to DM$, that gives rise to it.

Proposition 14 (Subdistributions)**.** *Inclusion of normalized distributions into subdistributions, $(\bullet)\colon MDX \to DMX$, defined on distributions by $(\perp)^\bullet = 1\,|\perp\rangle$ and $d^\bullet = d$ and on morphisms by $f^\bullet(x;y) = f(x;y)$, induces a monoidal distributive law. The monoidal Kleisli category of this distributive law is the category of subdistributions, subStoch.*

Proposition 15 (Renormalization)**.** *The following equation holds in the category of subdistributions.*

$$\mathsf{n}(f \mathbin{;} g) = \mathsf{n}(\mathsf{n}(f) \mathbin{;} g).$$

More generally, this equation holds up to almost-sure equivalence in any partial Markov category [DR23].

Proposition 16. *The normalization magmoid admits an action from the category of subdistributions,*

$$(\prec)\colon \mathsf{Norm}(X;Y) \times \mathsf{Subd}(Y;Z) \to \mathsf{Norm}(X;Z),$$

defined by $p \prec f = \mathsf{n}(p^{\bullet} ; f)$. *That is,* $p \prec \mathrm{id} = p$ *and* $p \prec (f ; g) = p \prec f \prec g$.

3.2 Partial distributions

Normalized channels are also the morphisms of another category, albeit with a different composition operation. The category of partial distributions, ParStoch, composes two normalized channels, $f \colon X \to MDY$ and $g \colon Y \to MDZ$, into the partial operation

$$(f \mathbin{⨾} g)(x; z) = \begin{cases} \sum_{v \in Y} f(x; v) \cdot g(v; z) & \text{when defined,} \\ \bot & \text{elsewhere.} \end{cases}$$

Proposition 17. *Failure of any non-total distribution, the natural transformation* $(-)^{\bot} \colon DM \to MD$*, defined by* $f^{\bot}(x) = f(x) \cdot [f(\bot) = 0]$ *induces a monoidal distributive law. The Kleisli category of this distributive law is the category of partial distributions,* ParStoch.

Partial distributions are the leading example of *quasi-Markov categories* [FGL+25, Sha25]. While the quasi-Markov category of distributions may play an important role when avoiding failure, let us agree that it does not address the problem of normalization: indeed, it marks with failure whenever a normalization problem is encountered.

4 Distributive Sesquilaws

4.1 Almost-distributive laws

Normalization satisfies all of the axioms of a distributive law, except for one. We must drop exactly one of the multiplicativity axioms of distributive laws to recover the structure of normalization.

An *almost distributive law* could be any candidate distributive law failing one of the axioms. More specifically, we could define S-multiplicative, S-unital, T-multiplicative, and T-unital almost distributive laws, respectively. In this terminology, a *weak distributive law* [Str09, GP20] would be a T-unital almost distributive law. For the rest of the text, however, let us focus on T-multiplicative almost distributive laws, and let us simply call them almost distributive laws.

Definition 18 (Almost distributive law)**.** An *almost distributive law* is a candidate distributive law failing the T-multiplicativity axiom.

Definition 19 (Monoidal almost-distributive law)**.** A *monoidal almost-distributive law* between two monoidal monads is an almost-distributive law whose underlying natural transformation is monoidal.

Proposition 20 (Kleisli magmoid of an almost-distributive law)**.** *Any almost-distributive law induces a magmoid. Any monoidal almost-distributive law induces a monoidal magmoid.*

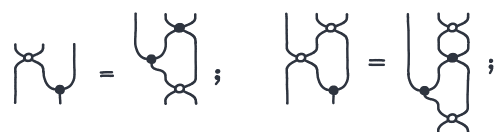
The monoidal almost-distributive law of normalization induces the Kleisli monoidal magmoid, Norm. It remains to study how it interacts with the actual distributive laws that describe subdistributions and partial distributions.

4.2 Distributive Sesquilaws

Normalization satisfies all the axioms for a distributive law $DM \to MD$ except for the D-multiplicativity axiom: as a result, its Kleisli construction is a non-associative category. However, normalization still satisfies $\mathsf{n}(\mathsf{n}(f) \mathbin{⨾} g) = \mathsf{n}(f \mathbin{⨾} g)$, if we reinterpret each non-failing element of MDX as a distribution in DX. This follows from the D-multiplicativity axiom holding *up-to-an-idempotent*: the distributive law of subdistributions, $MD \to DM$, is the partial inverse inducing this idempotent.

Distributive sesquilaws abstract this situation into a single equation. This single equation is exactly multiplicativity up to the idempotent determined by the two distributivity law candidates. Black-hole semantics form another distributive sesquilaw.

Definition 21 (Distributive sesquilaw)**.** A *distributive sesquilaw* between two monads, $(\bowtie, \asymp, S, T)$, consists of a distributive law $(\bowtie) \colon ST \to TS$ and a T-multiplication almost distributive law $(\asymp) \colon TS \to ST$ that satisfy any of the following two equivalent equations.



A distributive sesquilaw is enough to prove most of the facts of normalization we care about.

Proposition 22 (Renormalization)**.** *Any distributive sesquilaw,* $(\bowtie, \asymp, S, T)$*, induces an idempotent,* $(\asymp \mathbin{⨾} \bowtie) \colon TS \to TS$. *This idempotent is left-absorptive, meaning that the following equation holds.*

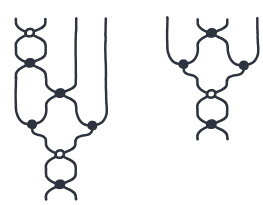
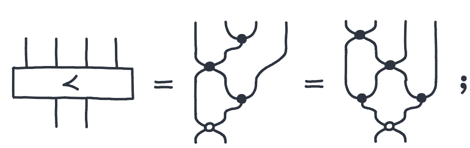


Figure 1. Renormalization equation.

Theorem 23. *Any distributive sesquilaw,* $(\bowtie, \asymp, S, T)$*, induces an action of* TS *into* ST*, defined as follows.*



This is a general phenomenon for distributive sesquilaws.

Theorem 24. *In the setting of a distributive sesquilaw,* $(\bowtie_{\bullet}, \bowtie_{\circ})$*, the Kleisli category of the distributive law acts on the Kleisli magmoid of the non-multiplicative distributive law.*

5 Guarded Choice and Tricocycloids

Finally, this section opens this framework to a very general notion of normalization: that described by tricocycloids [Gar18]. This includes the possibilistic case of relations and yields an important connection: distributions are to total relations what the normalization magmoid is to the category of relations.

5.1 Guarded choice from tricocycloids

Conisder an algebraic theory containing a family of guards: binary operations, $(\oplus_p)$, parameterized by some set of choices, $p \in H$. These choices may be a probability of choosing the left side — say $p = 0.4$ — or just a witness of non-determinism — say $p = \{L, R\}$. By convention, we avoid the two constant choices, as these can always be reduced, $x \oplus_L y = x$ and $x \oplus_R y = y$. Thus, while our probabilistic example requires $H = (0, 1)$; our possibilistic example only requires $H = \{*\}$.

Guards have an algebra of themselves [SFH+19]. At the very least, we expect them to allow rewriting by commutativity, and associativity. That is, there must exist a unary commutator $(^*)$ and two binary associators $(\bullet, \circ)$ satisfying the following equations.

1. $x \oplus_p y = y \oplus_{p^*} x$;
2. $(x \oplus_q y) \oplus_p z = x \oplus_{p \bullet q} (y \oplus_{p \circ q} z)$.

To allow rewriting in the opposite direction, the commutator must be bijective and the associators must be jointly bijective — meaning that each pair (p, q) must be of the form $(r \bullet s, r \circ s)$ for exactly one pair (r, s).

The previous equations allow us to arbitrarily move and regroup variables. For instance, for term depending non-trivially on $n + m$ variables, $t(x_1, ..., x_n, y_1, ..., y_m)$, we can always find terms $t_1(x_1, ..., x_n)$ and $t_2(y_1, ..., y_m)$ — depending on the first n and the last m variables — such that

$$t(x_1, ..., x_n, y_1, ..., y_m) = t_1(x_1, ..., x_n) \oplus_p t_2(y_1, ..., y_m),$$

for some guard p. In particular, for a non-trivial subdistribution $t(x_1, ..., x_n, \bot)$, we can always find a term $n(t)(x_1, ..., x_n)$ — its normalization — such that

$$t(x_1, ..., x_n, \bot) = n(t)(x_1, ..., x_n) \oplus_v k(\bot).$$

The guard of this normalization v is called the *validity.*

5.2 Confluence

Still, it remains to show that this rewriting system is confluent: it could be that different rewritings yield different guards; and we have not shown that these are unique. Fortunately, coherence for rewriting by symmetry and associativity is solved: it is well-known that MacLane's pentagon and hexagon equations [ML71] are all that is needed for symmetric coherence.

Remark 25. Consider rewriting according to MacLane's pentagon.

$$\begin{aligned}
&((x \oplus_r y) \oplus_q z) \oplus_p w && \mapsto \\
&(x \oplus_{q \bullet r} (y \oplus_{q \circ r} z)) \oplus_p w && \mapsto \\
&x \oplus_{p \bullet (q \bullet r)} ((y \oplus_{q \circ r} z) \oplus_{p \circ (q \bullet r)} w) && \mapsto \\
&x \oplus_{p \bullet (q \bullet r)} (y \oplus_{(p \circ (q \bullet r)) \bullet (q \circ r)} (z \oplus_{(p \circ (q \bullet r)) \circ (q \circ r)} w)).
\end{aligned}$$

$$\begin{aligned}
&((x \oplus_r y) \oplus_q z) \oplus_p w && \mapsto \\
&(x \oplus_r y) \oplus_{p \bullet q} (z \oplus_{p \circ q} w) && \mapsto \\
&x \oplus_{(p \bullet q) \bullet r} (y \oplus_{(p \bullet q) \circ r} (z \oplus_{p \circ q} w)).
\end{aligned}$$

From this critical pair, we obtain the following equations.

1. $p \bullet (q \bullet r) = (p \bullet q) \bullet r$;
2. $(p \bullet q) \circ r = (p \circ (q \bullet r)) \bullet (q \circ r)$;
3. $p \circ q = (p \circ (q \bullet r)) \circ (q \circ r)$.

Remark 26. Consider rewriting according to MacLane's hexagon.

$$\begin{aligned}
&(x \oplus_q y) \oplus_p z && \mapsto \\
&x \oplus_{p \bullet q} (y \oplus_{p \circ q} z) && \mapsto \\
&(y \oplus_{p \circ q} z) \oplus_{(p \bullet q)^*} x && \mapsto \\
&y \oplus_{(p \bullet q)^* \bullet (p \circ q)} (z \oplus_{(p \bullet q)^* \circ (p \circ q)} x).
\end{aligned}$$

$$\begin{aligned}
&(x \oplus_q y) \oplus_p z && \mapsto \\
&(y \oplus_{q^*} x) \oplus_p z && \mapsto \\
&y \oplus_{p \bullet q^*} (x \oplus_{p \circ q^*} z) && \mapsto \\
&y \oplus_{p \bullet q^*} (x \oplus_{(p \circ q^*)^*} z).
\end{aligned}$$

From this critical pair, we obtain the following equations.

1. $(p \bullet q)^* \bullet (p \circ q) = p \bullet q^*$;
2. $(p \bullet q)^* \circ (p \circ q) = (p \circ q^*)^*$.

These algebraic structures go back to Street, under the name of *symmetric tricocycloids* [Str98, Gar18].

Definition 27 (Symmetric tricocycloid)**.** A *symmetric tricocycloid*, $(H, \bullet, \circ, \bullet^{-1}, \circ^{-1}, ^*)$, is a set endowed with four binary operations and a unary operation satifying the following axioms.

1. $p^{**} = p$;
2. $(p \bullet q) \bullet^{-1} (p \circ q) = p$;
3. $(p \bullet q) \circ^{-1} (p \circ q) = q$;
4. $(p \bullet q)^* \bullet (p \circ q) = p \bullet q^*$;
5. $(p \bullet q)^* \circ (p \circ q) = (p \circ q^*)^*$;
6. $p \bullet (q \bullet r) = (p \bullet q) \bullet r$;
7. $(p \bullet q) \circ r = (p \circ (q \bullet r)) \bullet (q \circ r)$;
8. $p \circ q = (p \circ (q \bullet r)) \circ (q \circ r)$.

Example 28 ([Gar18]). The open interval, $(0, 1)$, is a symmetric tricocycloid with operations given by

$$p^* = 1 - p; \quad p \bullet q = p \cdot q; \quad p \circ q = \frac{p(1 - q)}{1 - pq}.$$

The singleton is the terminal symmetric tricocycloid.

Definition 29 (Lawvere theory induced by a tricocycloid)**.** The Lawvere theory induced by a symmetric tricocycloid $(H, \bullet, \circ, {}^*)$ contains a binary operation, $(\oplus_p) \in \mathcal{O}(2)$, for each element of the tricocycloid, $p \in H$, and satisfies the following equations:

1. idempotency, $x \oplus_p x = x$;
2. commutativity, $x \oplus_p y = y \oplus_{p^*} x$;
3. interchange, $(x \oplus_p y) \oplus_q (u \oplus_p v) = (x \oplus_q u) \oplus_p (y \oplus_q v)$;
4. and associativity, $(x \oplus_q y) \oplus_p z = x \oplus_{p \bullet q} (y \oplus_{p \circ q} z)$.

As a consequence, each symmetric tricocycloid induces a finitary monad, $D_H \colon \mathbf{Set} \to \mathbf{Set}$.

Lemma 30. *The Lawvere theory induced by a symmetric tricocycloid is commutative. Thus, the finitary monad of a symmetric tricocycloid is monoidal.*

Theorem 31. *The Kleisli category of the finitary monad $D_H \colon \mathbf{Set} \to \mathbf{Set}$ induced by a symmetric tricocycloid is a Markov category with conditionals.*

Theorem 32. *A finitary monad, $D_H \colon \mathbf{Set} \to \mathbf{Set}$, associated to a symmetric tricocycloid, H, admits a distributive law from the Maybe monad,* $\mathsf{sub} \colon MD_H \to D_H M$*, and two sections to it,* $\mathsf{quas} \colon D_H M \to MD_H$ *and* $\mathsf{norm} \colon D_H M \to MD_H$*, with the first one forming a distributive law and both forming a distributive sesquilaw.*

Proof sketch. It is well-known that the Maybe monad — and more generally, the Exception monad — distributes over any other monad [ZM20] with the obvious map $\mathsf{sub} \colon MD_H \to D_H M$. It remains to check the axiom of distributive sesquilaws for the second transformation. □

6 Stochastic relations

6.1 The Quest for Stochastic Relations

Parallel to the quest for a category of normalized kernels is the quest for a category of stochastic relations. The category of stochastic relations would have applications all across computer science — in process semantics, program logics, or process calculi — if only it were to exist.

(a) Distributions are stochastic non-empty relations. Naively, we could think that distributions are already a good notion of stochastic relations. Mapping each distribution to its support, $DX \to PX$, does define a monad morphism. However, something is off: this functor is not full, its image consists only of the non-empty relations. If we call N to the non-empty relations monad, only $\mathsf{supp} \colon DX \to NX$ is surjective.

(b) Subdistributions are not stochastic relations. To address the previous problem, we may add failure: adding failure to distributions results in subdistributions. Indeed, Panangaden [Pan99] defines *stochastic relations* to be substochastic Markov kernels.

Still, we may realize that subdistributions do not handle failure as relations do. Indeed, apart from the usual category of relations, there are two important categories of relations that get relatively less attention: subrelations (known as *May-Must relations*) and quasirelations (known as *Dijkstra relations*). It is easy to fall into these when trying to recover relations. We will show that subdistributions are stochastic subrelations, while quasidistributions are stochastic quasirelations.

Accepting the magmoid of stochastic relations. From this point of view, there seems to be no real analogue to the category of relations. That is, unless we are open to accept that the analogue is not a category but merely a magmoid. From this point of view, normalized kernels are stochastic relations. Let us justify this correspondence.

6.2 Support as a morphism of tricocycloids

Probability and possibility are related by taking supports. The support of a distribution consists of the subset of possible outcomes, forgetting about the specific probabilities of any of them.

Definition 33 (Support)**.** The *support* of a distribution defines a natural transformation, $\mathsf{supp} \colon D \to N$, given by $\mathsf{supp}(d) = \{x \mid d(x) > 0\}$. Note that, because it is a distribution, the set is always non-empty.

Support can be explained a morphism of tricocycloids, inducing a morphism of Lawvere theories, and inducing functors between the respective categories.

Proposition 34 (Terminal tricocycloid)**.** *The terminal symmetric tricocycloid is the singleton set.*

Let us compare two symmetric tricocycloids, the probabilistic tricocycloid — the interval $(0, 1)$ — and the possibilistic tricocycloid — the singleton $\{*\}$. The monad associated to the terminal tricocycloid is the monad of non-empty relations, which we denote by N; the monad associated to the interval tricocycloid is the finitary distribution monad. Supports of distributions compose as total relations; supports of subdistributions compose as may-must relations; supports of partial distributions compose by Dijkstra composition of relations; and normalized distributions compose as relations.

Definition 35 (May-Must)**.** The category mmRel of May-Must relations has as morphisms the channels $f \colon X \to NMY$, and composition $(\fatsemi)$ is defined by

$$(f \fatsemi g)(x; z) = \exists_{y \in Y} f(x; y) \wedge g(y; z),$$
$$(f \fatsemi g)(x; \bot) = f(x; \bot) \vee \exists_{y \in Y} f(x; y) \wedge g(y; \bot).$$

Proposition 36 (Support)**.** *Support extends to a functor*

$$\mathsf{supp} \colon \mathsf{subStoch} \to \mathsf{mmRel}.$$

Definition 37 (Dijkstra relations)**.** The category dkRel of Dijkstra relations has morphisms the channels $f \colon X \to MNY$

and composition ($\star$) is partially defined by

$$(f \star g)(x;z) = \exists_{y \in Y} f(x;y) \wedge g(y;z)$$

but resulting in failure whenever $\exists_{y \in Y} f(x;y) \wedge g(y, \bot)$.

Proposition 38 (Support)**.** *Support extends to a functor*

$$\mathrm{supp}\colon \mathsf{ParStoch} \to \mathsf{dkRel}.$$

Finally, the following extension justifies the magmoid of normalized stochastic kernels as a valid notion of stochastic relation.

Proposition 39. *Support extends to a functor of magmoids*

$$\mathrm{supp}\colon \mathsf{Norm} \to \mathsf{Rel}.$$

In summary, we have the following table relating (*i*) distributions to non-empty relations, (*ii*) subdistributions to May-Must relations, (*iii*) partial distributions to Dijkstra relations, and (*iv*) normalized distributions to relations.

Kleisli	Monad	Kleisli	Monad
Stoch	DIST	RelTot	P_{ne}
subStoch	$D_{\leq 1}$	mmRel	P_M
ParStoch	$D_{\bot}$	dkRel	P_D
Norm	D_N	Rel	P

Arguably, then, a valid notion of stochastic relations is that of normalized stochastic channels. Taking this point of view seriously, the magmoid of relations happens to be a category, but we should not expect the same from stochastic relations. Let us accept that stochastic relations form only a magmoid.

Acknowledgements

Elena Di Lavore and Mario Román were supported by the Advanced Research + Invention Agency (ARIA) Safeguarded AI Programme. Mario Román was supported by the Air Force Office of Scientific Research under award number FA9550-21-1-0038.

References

[Bec69] Jon Beck. Distributive laws. In *Seminar on triples and categorical homology theory*, pages 119–140. Springer, 1969.

[DR23] Elena Di Lavore and Mario Román. Evidential Decision Theory via Partial Markov Categories. In *2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–14. IEEE, 2023.

[FGL+25] Tobias Fritz, Tomáš Gonda, Antonio Lorenzin, Paolo Perrone, and Areeb Shah Mohammed. Empirical measures and strong laws of large numbers in categorical probability. *arXiv preprint arXiv:2503.21576*, 2025.

[Fri09] Tobias Fritz. Convex spaces i: Definition and examples. *arXiv preprint arXiv:0903.5522*, 2009.

[Gar18] Richard Garner. Abstract hypernormalisation, and normalisation-by-trace-evaluation for generative systems. *arXiv preprint arXiv:1811.02710*, 2018.

[GP20] Alexandre Goy and Daniela Petrisan. Combining weak distributive laws: Application to up-to techniques. *CoRR*, abs/2010.00811, 2020.

[Jac17] Bart Jacobs. Hyper normalisation and conditioning for discrete probability distributions. *Log. Methods Comput. Sci.*, 13(3), 2017.

[ML71] Saunders Mac Lane. *Categories for the Working Mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer Verlag, 1971.

[Pan99] Prakash Panangaden. The Category of Markov Kernels. *Electronic Notes in Theoretical Computer Science*, 22:171–187, January 1999.

[PTRSZ25] Robin Piedeleu, Mateo Torres-Ruiz, Alexandra Silva, and Fabio Zanasi. A complete axiomatisation of equivalence for discrete probabilistic programming. In *European Symposium on Programming*, pages 202–229. Springer, 2025.

[SFH+19] Steffen Smolka, Nate Foster, Justin Hsu, Tobias Kappé, Dexter Kozen, and Alexandra Silva. Guarded kleene algebra with tests: verification of uninterpreted programs in nearly linear time. *Proceedings of the ACM on Programming Languages*, 4(POPL):1–28, 2019.

[Sha25] Areeb Shah Mohammed. Partializations of markov categories, 2025.

[SS24] Dario Stein and Sam Staton. Probabilistic programming with exact conditions. *J. ACM*, 71(1):2:1–2:53, 2024.

[Str98] Ross Street. Fusion operators and cocycloids in monoidal categories. *Applied Categorical Structures*, 6(2):177–191, 1998.

[Str09] Ross Street. Weak distributive laws. *Theory and Applications of Categories*, 22:313–320, 2009.

[SW18] Ana Sokolova and Harald Woracek. Termination in convex sets of distributions. *Log. Methods Comput. Sci.*, 14(4), 2018.

[Ver17] Dominic Verdon. Coherence for braided and symmetric pseudomonoids. *CoRR*, abs/1705.09354, 2017.

[vS] Marilyn vos Savant. Parade 16: Ask Marilyn (Archived). https://web.archive.org/web/20130121183432/http://marilynvossavant.com/game-show-problem/. Accessed: 2013-01-21.

[ZM20] Maaike Zwart and Dan Marsden. No-go theorems for distributive laws. *CoRR*, abs/2003.12531, 2020.

A Proofs for Section 2 (Normalization)

Proposition 5. *The normalized distribution magmoid is not a category.*

Proof. Let us produce a concrete counterexample. Consider a coin flip, $f = 1/2\,|a\rangle + 1/2\,|b\rangle$, followed by a channel that marks it with two different failure probabilities $g(a) = 1/3\,|x\rangle + 2/3\,|z\rangle$ and $g(b) = 1/2\,|a\rangle + 1/2\,|b\rangle$, and followed by a channel that fails, $h(x) = |x\rangle$ and $h(y) = |y\rangle$, but $h(z) = 0$.

In this case, we have $(f \fatsemi g) \fatsemi h \neq f \fatsemi (g \fatsemi h)$, because of the following computation for the left-hand side,

$$\begin{aligned} &\overset{f}{\rightsquigarrow} && 1/2\,|a\rangle + 1/2\,|b\rangle \\ &\overset{g}{\rightsquigarrow} && 1/6\,|x\rangle + 2/6\,|z\rangle + 1/4\,|y\rangle + 1/4\,|z\rangle \\ &\overset{h}{\rightsquigarrow} && 2/5\,|x\rangle + 3/5\,|y\rangle\,. \end{aligned}$$

But we have that the right-hand side composition amounts to $(g \fatsemi h)(a) = 1\,|x\rangle$ and $(g \fatsemi h)(b) = 1\,|y\rangle$, and thus the result is $1/2\,|x\rangle + 1/2\,|y\rangle$. □

Proposition 7 (Associating morphisms form a subcategory)**.** *Associating morphisms of a magmoid form a category with the composition of the original magmoid.*

Proof. Let us first note that the identity is associating,

$$(f \mathbin{⨾} \mathrm{id}) \mathbin{⨾} g = f \mathbin{⨾} g = f \mathbin{⨾} (\mathrm{id} \mathbin{⨾} g).$$

And let us then note that, if two compatible morphisms, h_1 and h_2, are associating, then their composition, $h_1 \mathbin{⨾} h_2$, is also associating.

$$\begin{aligned}(f \mathbin{⨾} (h_1 \mathbin{⨾} h_2)) \mathbin{⨾} g &\overset{(i)}{=} ((f \mathbin{⨾} h_1) \mathbin{⨾} h_2) \mathbin{⨾} g\\ &\overset{(ii)}{=} (f \mathbin{⨾} h_1) \mathbin{⨾} (h_2 \mathbin{⨾} g)\\ &\overset{(iii)}{=} f \mathbin{⨾} (h_1 \mathbin{⨾} (h_2 \mathbin{⨾} g))\\ &\overset{(iv)}{=} f \mathbin{⨾} ((h_1 \mathbin{⨾} h_2) \mathbin{⨾} g).\end{aligned}$$

Where we have used *(i,iii)* that h_1 is associating; and *(ii,iv)* that h_2 is associating. □

Proposition 10 (Norm is monoidal)**.** *The normalized distribution magmoid is monoidal with the cartesian product of sets and the following partial product of morphisms.*

$$(f_1 \otimes f_2)(x_1, x_2; y_1, y_2) = f_1(x_1; y_1) \cdot f_2(x_2; y_2).$$

Proof. Most equations are direct to check; let us check that it satisfies the interchange law.

$$\begin{aligned}&((f_1 \otimes f_2) \mathbin{⨾} (g_1 \otimes g_2))(x_1, x_2; z_1, z_2) \quad =\\ &\frac{\sum_{y_1,y_2} f_1(x_1; y_1) \cdot f_2(x_2; y_2) \cdot g_1(y_1; z_1) \cdot g_2(y_2; z_2)}{\sum_{y_1,y_2,z_1,z_2} f_1(x_1; y_1) \cdot f_2(x_2; y_2) \cdot g_1(y_1; z_1) \cdot g_2(y_2; z_2)} \quad =\\ &\frac{\sum_{y_1} f_1(x_1; y_1) \cdot g_1(y_1; z_1)}{\sum_{y_1,z_1} f_1(x_1; y_1) \cdot g_1(y_1; z_1)} \cdot \frac{\sum_{y_2} f_2(x_2; y_2) \cdot g_2(y_2; z_2)}{\sum_{y_2,z_2} f_2(x_2; y_2) \cdot g_2(x_2; z_2)} \quad =\\ &(f_1 \mathbin{⨾} g_1)(x_1; z_1) \cdot (f_2 \mathbin{⨾} g_2)(x_2; z_2).\end{aligned}$$

We use (*i*) the definition of the composition and tensor, (*ii*) distributivity of products over sums, and (*iii*) the definition of composition and tensor, again. □

B Proofs for Section 3 (Distributive Laws)

Proposition 15 (Renormalization)**.** *The following equation holds in the category of subdistributions.*

$$\mathsf{n}(f \,;\, g) = \mathsf{n}(\mathsf{n}(f) \,;\, g).$$

Proposition 16. *The normalization magmoid admits an action from the category of subdistributions,*

$$(\prec) \colon \mathsf{Norm}(X; Y) \times \mathsf{Subd}(Y; Z) \to \mathsf{Norm}(X; Z),$$

defined by $p \prec f = \mathsf{n}(p^{\bullet}; f)$*. That is,* $p \prec \mathrm{id} = p$ *and* $p \prec (f; g) = p \prec f \prec g$*.*

Proof. The result follows from the application of Theorem 15.

$$p \prec (f \mathbin{⨾} g) = \mathsf{n}(p^{\circ} \mathbin{⨾} f \mathbin{⨾} g) = \mathsf{n}(\mathsf{n}(p^{\circ} \mathbin{⨾} f) \mathbin{⨾} g) = \mathsf{n}(p^{\circ} \mathbin{⨾} f) \prec g = p \prec f \prec g.$$

□

C Proofs for Section 4 (Distributive Sesquilaws)

Proposition 22 (Renormalization)**.** *Any distributive sesquilaw,* (✕•, ✕∘, S, T)*, induces an idempotent,* (✕∘ ⨾ ✕•)$\colon TS \to TS$*. This idempotent is left-absorptive, meaning that the following equation holds.*

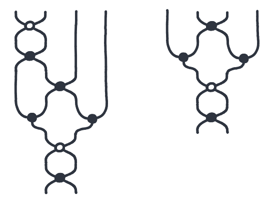
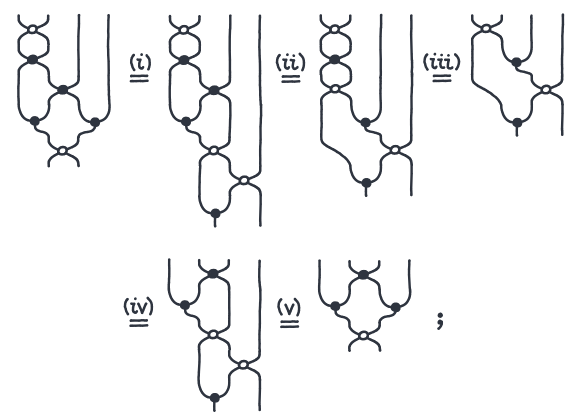


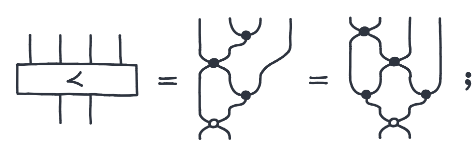
Figure 2. Renormalization equation.

Proof. Let us prove a slightly stronger equation where we omit the lsat composition with the distributive law (✕•). In Section C, we use *(i)* the multiplicativity axiom, *(ii)* the distributive sesquilaw equation, *(iii)* that distributive sesquilaws are inverses, *(iv)* the distributive sesquilaw equation, *(v)* the multiplicativity axiom.



This concludes the proof. □

Theorem 23. *Any distributive sesquilaw,* (✕•, ✕∘, S, T)*, induces an action of* TS *into* ST*, defined as follows.*



Proof. We reason by string diagrams (Figure 3). We use *(i)* the multiplicativity axiom, *(ii)* that distributive swaps are inverses, *(iii)* the distributive sesquilaw equation, *(iv)* the multiplicativity axiom, *(v)* the multiplicativity of the distributive law, *(vi)* associativity of the monad, and *(vii,viii)* the multiplicativity of the distributive law.

This concludes the proof. □

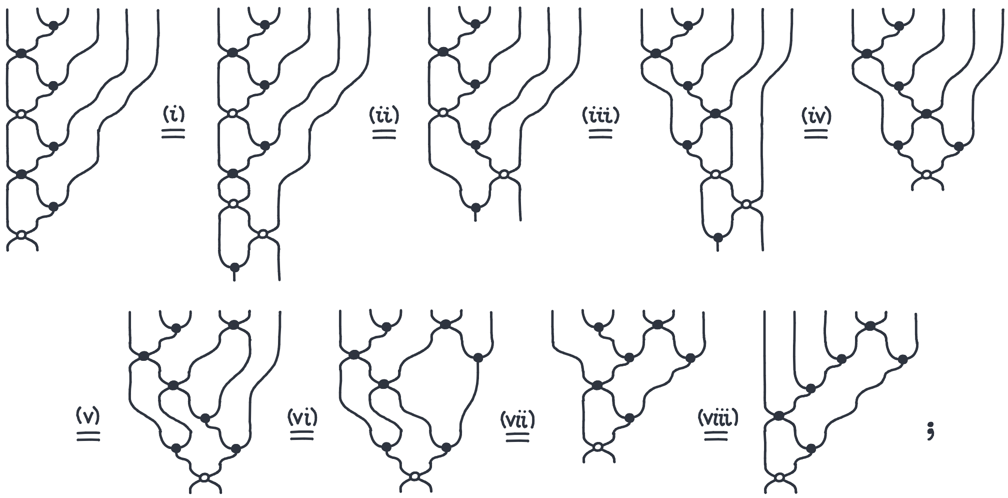


Figure 3. Proof of the multiplicativity of the action induced by a distributive sesquilaw.