Accepting Normalization via Markov Magmoids

Elena Di Lavore
Mario Román

Abstract

Normalization is not a distributive law, but just an almost-distributive law that is a section to an actual distributive law. We introduce distributive swaps to describe this situation and derive synthetically multiple facts about normalization. We then introduce Markov magmoids, a non-associative variant of Markov categories with conditionals, having as the lead example the category of normalized channels.

Keywords: Category theory, categorical semantics.

1 Introduction

Normalization is difficult to accept in category theory. 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]. Accepting normalization requires a change of perspective: we must accept normalization for the structure it has, not the structure it fails to have.

And this 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 an abstraction of distributive laws.

1.1 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 $\sum_{x'\in X} f(x') = 0$.

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 be monoidal. The tragedy is that normalization fails to be a distributive law, and this potential Kleisli category is instead a Kleisli magmoid.

1.2 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

$$(\fatsemi)\colon \mathbb{A}(X;Y)\times\mathbb{A}(Y;Z) \to \mathbb{A}(X;Z), \text{ and}$$
$$\mathrm{id}\colon \mathbb{A}(X;X);$$

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

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*

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

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

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

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

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.*

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.*

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, every monoidal magmoid is equivalent to a strict one.

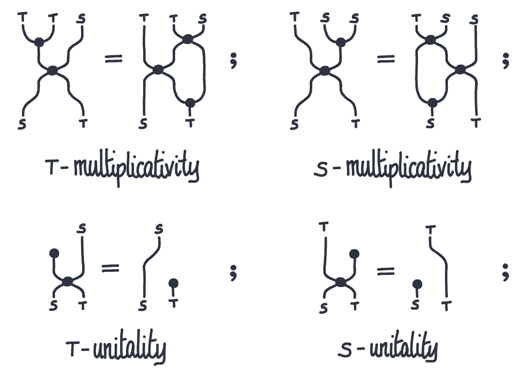
Proposition 10. *The normalized distribution magmoid is monoidal with the cartesian product of sets and the following partial product of morphisms.*

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

2 Distributive Laws

Distributive laws [Bec69], their uses and failures [ZM20], are all well-known. Let us quickly recap. 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.

Theorem 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 $S \circ T$.*

2.1 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 by $f^\bullet(x;y) = f(x;y)$, induces a monoidal distributive law. The Kleisli category of this distributive law is the category of subdistributions.*

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

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

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

2.2 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

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

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, Moh25]. While the quasi-Markov category of distributions will play an important role later on, let us agree that it does not address the problem of normalization either: instead, it marks with failure whenever a normalization problem is encountered.

3 Distributive Swaps

3.1 Almost-distributive laws

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

Definition 18 (Almost distributive law)**.** An *almost distributive law* is a candidate distributive law failing one of the axioms. More specifically, we define *Sm-almost distributive laws*, *Su-almost distributive laws*, *Tm-almost distributive laws*, and *Tu-almost distributive laws*, respectively.

Remark 19. A *weak distributive law* [Str09, GP20] is a *Tu*-almost distributive law in this terminology. During the rest of the text, we focus on *Tm*-almost distributive laws, and we simply call these almost-distributive laws.

Definition 20 (Monoidal almost-distributive law)**.** A *monoidal almost-distributive law* between two monoidal monads is an almost-distributive law whose underlying natural transformation is monoidal.

The monoidal almost-distributive law of normalization induces the Kleisli monoidal magmoid, Norm.

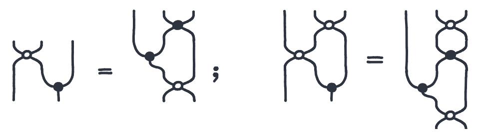
Proposition 21 (Kleisli magmoid of an almost-distributive law)**.** *Any almost-distributive law induces a magmoid. Any monoidal almost-distributive law induces a monoidal magmoid.*

3.2 Distributive Swaps

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 $\mathsf{n}(\mathsf{n}(f) \fatsemi g) = \mathsf{n}(f \fatsemi g)$, if we reinterpret each non-failing element of MDX as a distribution in DX. This follows from the D-multiplicativity rule holding up to an idempotent: the distributive law of subdistributions, $MD \to DM$, is a partial inverse. The situation follows form being a partial inverse and a distributive law, and it also holds true for the "black-hole" or "squashing" distributive law.

Distributive swaps abstract this situation into a single equation. This single equation is exactly multiplicativity up to the idempotent determined by the two distributivity law candidates.

Definition 22 (Distributive swap)**.** A *distributive swap* between two monads, (✖, ⨯, S, T), consists of a distributive law (✖): $ST \to TS$ and a T-multiplication almost distributive law (⨯): $TS \to ST$ that satisfy any of the following two equivalent equations.



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

Proposition 23 (Renormalization)**.** *Any distributive swap,* (✖, ⨯, S, T)*, induces an idempotent,* (⨯ ⨟ ✖)$\colon TS \to TS$*. This idempotent is left-absorptive, meaning that the following equation holds.*

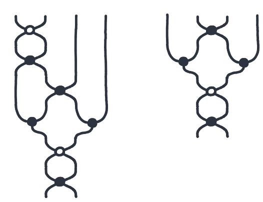
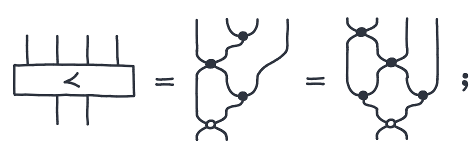


Figure 1. Renormalization equation.

Theorem 24. *Any distributive swap,* (✖, ⨯, S, T)*, induces an action of* TS *into* ST*, defined as follows.*



This is a general phenomenon for distributive swaps.

Theorem 25. *In the setting of a distributive swap,* (✖, ⨯)*, the Kleisli category of the distributive law acts on the Kleisli magmoid of the non-multiplicative distributive law.*

4 Related work

Every tricocycloid [Gar18] gives rise to a distributive swap. Morphisms of tricocycloids induce functors between the related Markov constructions. In particular, the singleton terminal tricocycloid induces the categories of non-empty relations, may-must relations, Dijkstra relations, and relations; the universal map to the terminal tricocycloid is the support map from distributions, subdistributions, partial distributions, and normalized distributions.

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.

[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.

[Moh25] Areeb Shah Mohammed. Partializations of markov categories, 2025.

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

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

A Proofs for Section 1 (Introduction)

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,

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

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

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

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

B Proofs for Section 2 (Distributive Laws)

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

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

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

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

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

□

C Proofs for Section 3 (Distributive Swaps)

Proposition 23 (Renormalization)**.** *Any distributive swap, $(\times, \times, S, T)$, induces an idempotent, $(\times \fatsemi \times)\colon TS \to TS$. This idempotent is left-absorptive, meaning that the following equation holds.*

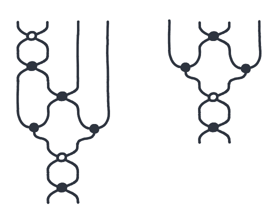


Figure 2. Renormalization equation.

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 swap equation, *(iii)* that distributive swaps are inverses, *(iv)* the distributive swap equation, *(v)* the multiplicativity axiom.

This concludes the proof. □

Theorem 24. *Any distributive swap,* (✖, ✕, S, T)*, induces an action of* TS *into* ST*, defined as follows.*

Proof. We reason by string diagrams (Section C). We use *(i)* the multiplicativity axiom, *(ii)* that distributive swaps are inverses, *(iii)* the distributive swap 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. □

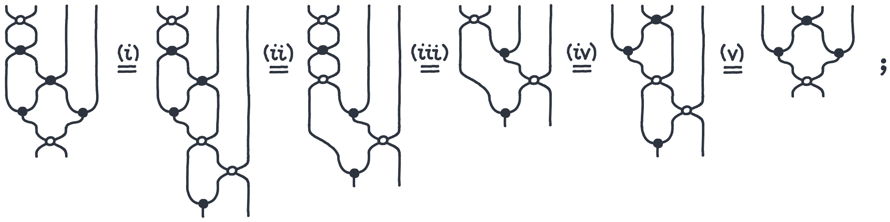


Figure 3. Proof of the abstract renormalization equation.

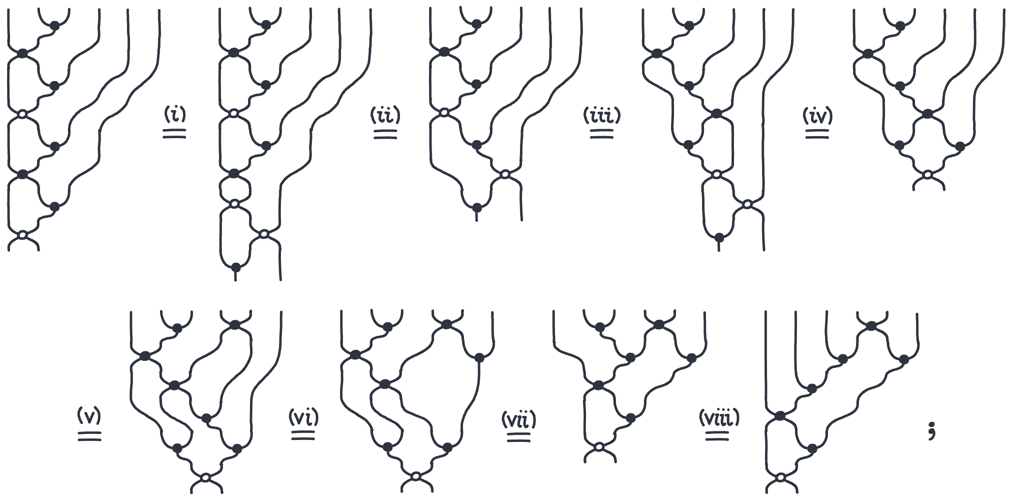


Figure 4. Proof of the multiplicativity of the action induced by a distributive swap.