The Magmoid of Normalized Stochastic Kernels

EXTENDED ABSTRACT

Elena Di Lavore
Tallinn University of Technology
Estonia

Mario Román
Tallinn University of Technology
Estonia

Márk Széles
Radboud University Nijmegen
The Netherlands

Abstract

Normalization, D(X + 1) → D(X) + 1, is almost a distributive law; but because one of the distributive law axioms only holds up-to-idempotent, it yields a non-associative composition of normalized kernels. We introduce the Markov magmoid of normalized stochastic kernels: a normalized-by-construction semantics for probabilistic inference, unifying exact Bayesian observations and interventions as two parenthesizations of the same composite. Front-door and back-door criteria follow from the axioms of Markov magmoids; we implement these with non-associative monadic notation.

This is an extended abstract for "The Magmoid of Normalized Stochastic Kernels" [DRS26a], which contains full details.

1 Normalized kernels are not associative

Imagine solving the famous *Monty Hall problem* [Sel75]: in a game show, you choose from three closed doors for a chance of winning the prize behind one of them; however, after choosing, the host opens one of the empty doors and — with two closed doors standing — invites you to switch your guess. *Should you switch?*

Let us compute. We *(i)* consider a prior uniform probability that a prize is behind any of three doors (L, M, R); then, *(ii)* after choosing, say, the middle door, the host will randomly and uniformly open a door (L, M, R) that cannot be neither the chosen door nor the one with the prize; *(iii)* we then *observe* that the host opens, e.g., the left door (L); and, upon this, *(iv)* we renormalize the remaining probabilities and conclude we should switch to the right door (R): it doubles our chances of getting the prize.

$$\begin{array}{ll}
\textit{(i)} & \tfrac{1}{3}\,|L\rangle + \tfrac{1}{3}\,|M\rangle + \tfrac{1}{3}\,|R\rangle \\
\textit{(ii)} & \tfrac{1}{3}\,|L\rangle|R\rangle + \tfrac{1}{6}\,|M\rangle|L\rangle + \tfrac{1}{6}\,|M\rangle|R\rangle + \tfrac{1}{3}\,|R\rangle|L\rangle \\
\textit{(iii)} & \cancel{\tfrac{1}{3}\,|L\rangle|R\rangle} + \tfrac{1}{6}\,|M\rangle|L\rangle + \cancel{\tfrac{1}{6}\,|M\rangle|R\rangle} + \tfrac{1}{3}\,|R\rangle|L\rangle \\
\textit{(iv)} & \tfrac{1}{3}\,|M\rangle|L\rangle + \tfrac{2}{3}\,|R\rangle|L\rangle\,.
\end{array}$$

The phenomenon we seek to study occurs at the last two steps. We *(iii)* obtain something that is not a full distribution, but only a subdistribution; and then *(iv)* we multiply by a constant — by 2, in this example — to obtain again a distribution. This interplay between normalized and unnormalized distributions leads one to pick substochastic kernels for probabilistic semantics [Pan99, BDGS16]: functions $X \to \mathbf{DM}Y$, for $\mathbf{D}$ the distribution monad and $\mathbf{M}$ the maybe monad. The alternative is to use unnormalized distributions [Koz81, Sta17, WCGC18, EPT17].

However, we may wish to avoid substochasticity: if the last two steps were compressed into one, subdistributions would never appear. Instead, we would work with normalized kernels, $X \to \mathbf{MD}Y$, yielding either nothing or a full distribution. May we compose normalized kernels? Is that what we just did? From recent work by Sokolova and Woracek, we are forced to admit that there is no distributive law giving rise to normalized kernels.

Corollary 1.1 (from [SW18, Theorem 5.3])**.** *Black-hole semantics determines the only distributive law between the distribution monad and the maybe monad,* $\mathbf{DM} \to \mathbf{MD}$.

Instead, we claim that normalized kernel composition is naturally non-associative. Indeed, the *Monty Hall problem* admits another parenthesization. Imagine we consider everything since the action of the host as a parenthesized subproblem and we *(iv)* normalize internally before *(v)* normalizing globally.

$$\begin{array}{ll}
\textit{(i)} & \tfrac{1}{3}\,|L\rangle + \tfrac{1}{3}\,|M\rangle + \tfrac{1}{3}\,|R\rangle \\
\textit{(ii)} & \tfrac{1}{3}\,|L\rangle|R\rangle + \tfrac{1}{3}\,|M\rangle\left(\tfrac{1}{2}\,|L\rangle + \tfrac{1}{2}\,|R\rangle\right) + \tfrac{1}{3}\,|R\rangle|L\rangle \\
\textit{(iii)} & \tfrac{1}{3}\,|L\rangle\cancel{|R\rangle} + \tfrac{1}{3}\,|M\rangle\left(\tfrac{1}{2}\,|L\rangle + \cancel{\tfrac{1}{2}\,|R\rangle}\right) + \tfrac{1}{3}\,|R\rangle|L\rangle \\
\textit{(iv)} & \tfrac{1}{3}\,|M\rangle|L\rangle + \tfrac{1}{3}\,|R\rangle|L\rangle \\
\textit{(v)} & \tfrac{1}{2}\,|M\rangle|L\rangle + \tfrac{1}{2}\,|R\rangle|L\rangle\,.
\end{array}$$

In this case, *(iii)* we *intervene* to force the host to open the left door — say, the game show halts otherwise. Because it is forced, the host's decision stops carrying any inferential information: we are equally likely to see it no matter where the prize is.

The interpretation of the Monty Hall problem has been famously controversial [Sel75, vS]. Arguably, the difference of interpretation is clearer knowing that normalized kernel composition is not associative: *when to normalize* does change the result.

2 Normalization

Normalized distributions are, equivalently, subdistributions adding up to exactly 0 or 1,

$$\mathbf{MD}X \cong \left\{ d \in \mathbf{DM}X \;\middle|\; \sum_{x \in X} d(x) = 0 \text{ or } \sum_{x \in X} d(x) = 1 \right\}.$$

Under this interpretation, zero means failure: let us convene that division by zero yields zero for the next definition.

Definition 2.1 (Normalization)**.** *Normalization* is the natural transformation, $\mathsf{n}_X \colon \mathbf{DM}X \to \mathbf{MD}X$, defined by

$$\mathsf{n}(f)(x) = \left[\frac{f(x)}{\sum_{x' \in X} f(x')} \right].$$

Proposition 2.2 (Non-associative category of normalized kernels)**.** *Normalized kernels between sets, $X \to \mathbf{MD}Y$, form a non-associative category,* **Norm***, where composition of two morphisms, $f \colon X \to \mathbf{MD}Y$ and $g \colon Y \to \mathbf{MD}Z$, is defined by*

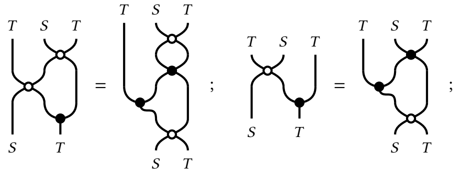
$$(f \fatsemi g)(x; z) = \left[\frac{\sum_{v \in Y} f(x; v) \cdot g(v; z)}{\sum_{v \in Y} \sum_{w \in Z} f(x; v) \cdot g(v; w)} \right].$$

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

3 Sesquilaws

We introduce sesquilaws to abstract this situation: a sesquilaw is one-and-a-half distributive laws. Well-behaved sesquilaws over copy-discard categories induce Markov magmoids.

Definition 3.1 (Sesquilaw)**.** A *sesquilaw*, $(S, T,$ ✖, ⤫$)$, between two monads, consists of a distributive law (✖): $ST \to TS$ and an almost distributive law (⤫): $TS \to ST$ — meaning a distributive law failing the T-multiplicativity axiom. These must form a section-retraction pair, (✖) ⨟ (⤫) = id, and satisfy any of the following two equivalent equations. A sesquilaw is *monoidal* whenever its laws are.



Proposition 3.2. *A copy-discard sesquilaw, (S, T, m, n), is a monoidal sesquilaw over a cartesian monoidal category, $(\mathbb{C}, \times, 1, \delta, \varepsilon)$, between an affine monad, $(T, \mu^T, \eta^T, u^T, v^T)$, and a relevant monad, $(S, \mu^S, \eta^S, u^S, v^S)$. Any copy-discard sesquilaw induces a quasitotal magmoid.*

Theorem 3.3. Norm *is a Markov magmoid: a left-relevant*[1] *quasitotal magmoid admitting conditionals.*

3.1 Example: continuous probability

Normalization, in the continuous case, is also a section to the distributive law that induces Panangaden's variant of the Giry monad.

Definition 3.4 (Panangaden monad, [Pan99, §3])**.** Standard Borel spaces admit the Giry monad $\mathcal{D} \colon \textbf{BorelMes} \to \textbf{BorelMes}$; they have coproducts and a Maybe monad, $\mathcal{M} \colon \textbf{BorelMes} \to \textbf{BorelMes}$. There exists a distributive law, $(-)^{\bullet}_{X} \colon \mathcal{MD}X \to \mathcal{DM}X$, yielding a monad structure on $\mathcal{DM}$ (c.f. [DR23]); its Kleisli category is **BorelSubstoch**.

Theorem 3.5. *Normalization,* $\mathsf{N}(-)_X \colon \mathcal{DM}X \to \mathcal{MD}X$, *induces a section to the substochastic distributive law of Definition 3.4; they moreover form a copy-discard sesquilaw.*

Corollary 3.6 (Magmoid of normalized kernels)**.** *Normalized kernels between standard Borel spaces, $X \to \mathcal{MD}Y$, form a monoidal non-associative category,* **BorelNorm**, *where composition of two kernels, $f \colon X \to \mathcal{MD}Y$ and $g \colon Y \to \mathcal{MD}Z$, is defined by*

$$(f \mathbin{⨟} g)(x; U) = \left[\frac{\int_{y \in Y} g(y; U) \cdot f(x; \mathrm{d}y)}{\int_{y \in Y} g(y; Z) \cdot f(x; \mathrm{d}y)} \right].$$

In other words, if we consider the associated substochastic kernels, $f^{\bullet} \colon X \to \mathcal{DM}Y$ and $g^{\bullet} \colon Y \to \mathcal{DM}Z$, it is the normalization of their composition, $f \mathbin{⨟} g = \mathsf{N}(f^{\bullet} \mathbin{;} g^{\bullet})$. The tensor of normalized kernels coincides with the usual one: we define $(f_1 \otimes f_2)((x_1, x_2); U)$ as

$$\int_{y_2 \in Y_2} \left(\int_{y_1 \in Y_1} \xi_U(y_1, y_2) \cdot f(x_1; \mathrm{d}y_2) \right) \cdot f_2(x_2; \mathrm{d}y_2),$$

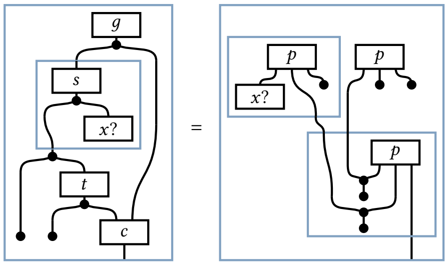
Theorem 3.7. **BorelNorm** *is a Markov magmoid.*

4 Programming in Markov magmoids

As an illustration, let us apply the more refined algebraic structure of Markov magmoids to a classical example from Pearl's work on causality.

Example 4.1 (Smoking causes cancer, [Pea88, DRS26b]). We know that smoking (s) may influence lung tar deposits (t), which may in turn influence the incidence of cancer (c); but imagine we still suspect that a gene (g) may be influencing both this incidence and a predisposition to smoke: may we quantify these causal influences? In principle, we would need data extracted from an interventional study forcing patients to smoke — which, previsibly, we want to avoid.

Instead, Markov magmoids can be used to derive a synthetic version of the front door adjustment formula [Pea09] in terms of string diagrams. String diagrams for magmoids, fortunately, resemble string diagrams with *normalization boxes* [LT23, JSS25, DRS25].



In other words, an intervention can be rewritten as a composition, in the Markov magmoid, of the observational joint data (p).

Using non-associative do-notation — which we introduce — the previous diagram gets translated to a program [DRS26b] estimating the effects of a causal intervention from observational data.

```
lDo Norm
   z <- (lDo Norm
             (list xi z yi) <- p
             '() <- (observe xq xi)
             return z)
   x <- (lDo Norm
             (list x zi yi) <- p
             return x)
   y <- (lDo Norm
             (list xi zi y) <- p
             '() <- (observe x xi)
             '() <- (observe z zi)
             return y)
  return y
```

From this perspective, causal inference is the resolution of equations in discrete Markov magmoids. Alternatively to most probabilistic programming literature, we do not use a normalization operator, but the ability to modulate associativity of a Markov magmoid.

Acknowledgements

We thank Soichiro Fujii, Bart Jacobs, Jack Liell-Cock, Paolo Perrone, Seo Jin Park, Zev Shirazi, Ana Sokolova, Owen Lynch, Sam Staton, Yun Chen Tsai, and Ruben van Belle for discussion on earlier drafts.

---

[1] Left-relevance is a property needed for the string diagrams of Section 4; please refer to the full paper for details [DRS26a].

Full references are provided for the main text: "The Magmoid of Normalized Stochastic Kernels" [DRS26a].

References

[BDGS16] Johannes Borgström, Ugo Dal Lago, Andrew D. Gordon, and Marcin Szymczak. A lambda-calculus foundation for universal probabilistic programming. In Jacques Garrigue, Gabriele Keller, and Eijiro Sumii, editors, *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming, ICFP 2016, Nara, Japan, September 18-22, 2016*, pages 33–46. ACM, 2016. doi:10.1145/2951913.2951942.

[DR23] Elena Di Lavore and Mario Román. Evidential decision theory via partial Markov categories. In *38th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2023, Boston, MA, USA, June 26-29, 2023*, pages 1–14. IEEE, 2023. doi:10.1109/LICS56636.2023.10175776.

[DRS25] Elena Di Lavore, Mario Román, and Paweł Sobociński. Partial Markov categories. *arXiv preprint*, 2025. doi:10.48550/arXiv.2502.03477.

[DRS26a] Elena Di Lavore, Mario Román, and Márk Széles. The Magmoid of Normalized Stochastic Kernels, 2026, 2510.01131. URL https://arxiv.org/abs/2510.01131.

[DRS26b] Elena Di Lavore, Mario Román, and Márk Széles. Racket v9.0 Implementation of the Markov Magmoid of Normalized Stochastic Kernels, 2026. doi:10.5281/zenodo.18348005.

[EPT17] Thomas Ehrhard, Michele Pagani, and Christine Tasson. Measurable cones and stable, measurable functions: a model for probabilistic higher-order programming. *Proceedings of the ACM on Programming Languages*, 2(POPL):1–28, 2017. doi:10.1145/3158147.

[JSS25] Bart Jacobs, Márk Széles, and Dario Stein. Compositional inference for Bayesian networks and causality. In *Mathematical Foundations of Programming Semantics*, 2025.

[Koz81] Dexter Kozen. Semantics of probabilistic programs. *J. Comput. Syst. Sci.*, 22(3):328–350, 1981. doi:10.1016/0022-0000(81)90036-2.

[LT23] Robin Lorenz and Sean Tull. Causal models in string diagrams, 2023, 2304.07638.

[Pan99] Prakash Panangaden. The Category of Markov Kernels. *Electronic Notes in Theoretical Computer Science*, 22:171–187, January 1999. doi:10.1016/S1571-0661(05)80602-4.

[Pea88] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan kaufmann, 1988.

[Pea09] Judea Pearl. *Causality.* Cambridge University Press, 2009.

[Sel75] Steve Selvin. Letters to the editor. *The American Statistician*, 29(1):67–71, 1975. doi:10.1080/00031305.1975.10479121.

[Sta17] Sam Staton. Commutative semantics for probabilistic programming. In *European Symposium on Programming*, pages 855–879. Springer, 2017.

[SW18] Ana Sokolova and Harald Woracek. Termination in convex sets of distributions. *Log. Methods Comput. Sci.*, 14(4), 2018. doi:10.23638/LMCS-14(4:17)2018.

[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.

[WCGC18] Mitchell Wand, Ryan Culpepper, Theophilos Giannakopoulos, and Andrew Cobb. Contextual equivalence for a probabilistic language with continuous random variables and recursion. *Proc. ACM Program. Lang.*, 2(ICFP):87:1–87:30, 2018. doi:10.1145/3236782.

Rights statement