HAL open science

Borel Kernels and their Approximation, Categorically

Fredrik Dahlqvist, Alexandra Silva, Vincent Danos, Ilias Garnier

To cite this version:

Fredrik Dahlqvist, Alexandra Silva, Vincent Danos, Ilias Garnier. Borel Kernels and their Approximation, Categorically. Electronic Notes in Theoretical Computer Science, 2018, 341, pp.91-119. 10.1016/j.entcs.2018.11.006 . hal-01976416

HAL Id: hal-01976416

https://hal.science/hal-01976416v1

Submitted on 10 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Borel Kernels and their Approximation, Categorically

Fredrik Dahlqvist
University College London
f.dahlqvist@ucl.ac.uk

Vincent Danos
Ecole Normale Supérieure Paris
Vincent.Danos@ens.fr

Ilias Garnier
Ecole Normale Supérieure Paris
Ilias.Garnier@ens.fr

Alexandra Silva
University College London
alexandra.silva@ucl.ac.uk

arXiv:1803.02651v2 [cs.LO] 8 Mar 2018

Abstract

This paper introduces a categorical framework to study the exact and approximate semantics of probabilistic programs. We construct a dagger symmetric monoidal category of Borel kernels where the dagger-structure is given by Bayesian inversion. We show functorial bridges between this category and categories of Banach lattices which formalize the move from kernel-based semantics to predicate transformer (backward) or state transformer (forward) semantics. These bridges are related by natural transformations, and we show in particular that the Radon-Nikodym and Riesz representation theorems - two pillars of probability theory - define natural transformations.

With the mathematical infrastructure in place, we present a generic and endogenous approach to approximating kernels on standard Borel spaces which exploits the involutive structure of our category of kernels. The approximation can be formulated in several equivalent ways by using the functorial bridges and natural transformations described above. Finally, we show that for sensible discretization schemes, every Borel kernel can be approximated by kernels on finite spaces, and that these approximations converge for a natural choice of topology.

We illustrate the theory by showing two examples of how approximation can effectively be used in practice: Bayesian inference and the Kleene $*$ operation of ProbNetKAT.

Keywords Probabilistic programming, probabilistic semantics, Markov process, Bayesian inference, approximation

1 Introduction

Finding a good category in which to study probabilistic programs is a subject of active research [6, 19, 23, 24]. In this paper we present a dagger symmetric monoidal category of kernels whose dagger-structure is given by Bayesian inversion. The advantages of this new category are two-fold.

Firstly, the most important new construct introduced by probabilistic programming, *viz. Bayesian inversion*, is interpreted completely straightforwardly by the †-operation which is native to our category. In particular we never leave the world of kernels and we therefore do not require any normalization construct. Consider for example the following simple Bayesian inference problem in Anglican ([26])

```
(defquery example
  (let [x (sample (normal 0 1))]
    (observe (normal x 1) 0.5)
    (> x 1)))
```

The semantics of this program is build easily and compositionally in our category:

- The second line builds a Borel space equipped with a normally distributed probability measure – an object $(\mathbb{R}, \mu)$ of our category.
- The `(normal x 1)` instruction builds a Borel kernel – a morphism $f : (\mathbb{R}, \mu) \to (\mathbb{R}, \nu)$ in our category.
- The `observe` statement builds the Bayesian inverse of the kernel – the morphism $f^\dagger : (\mathbb{R}, \nu) \to (\mathbb{R}, \mu)$ in our †-category.
- Finally, the kernel $f^\dagger$ is evaluated, i.e. the denotation of the program above is $f^\dagger(0.5)(]1, \infty[)$.

The functoriality of † ensures compositionality.

Second, since Bayesian inference problems are in general very hard to compute (although the one given above has an analytical solution), it makes sense to seek approximate solutions, i.e. *approximate denotations* to probabilistic programs. As we will show, our category of kernels comes equipped with a generic and endogenous approximating scheme which relies on its involutive structure and on the structure of standard Borel spaces. Moreover, this approximation scheme can be shown to converge for any choice of kernel for a natural choice of topology.

Main contributions.

1. We build a category **Krn** of Borel kernels (§2) and we show how two kernels which agree almost everywhere can be identified under a categorical quotient operation. This technical construction is what allows us to define *Bayesian inversion* as an involutive functor, denoted †. This is a key technical improvement on [6] where the †-structure[1] was hinted at but was not functorial. We show that **Krn** is a dagger symmetric monoidal category.

[1] Suggested to us by Chris Heunen.

Pre-print, March 2018

2. We introduce the category $\mathbf{BL}_\sigma$ of Banach lattices and σ-order continuous positive operators as well as the Köthe dual functor $(-)^\sigma : \mathbf{BL}_\sigma^{\mathrm{op}} \to \mathbf{BL}_\sigma$ (§3). These will play a central role in studying convergence of our approximation schemes.
3. We provide the first[2] categorical understanding of the Radon-Nikodym and the Riesz representation theorems. These arise as natural transformations between two functors relating kernels and Banach lattices (§4).
4. We show how the †-structure of **Krn** can be exploited to approximate kernels by averaging (§5). Due to an important structural feature of **Krn** (Th. 1) every kernel in **Krn** can be approximated by kernels on *finite* spaces.
5. We show a natural class of approximations schemes where the sequence of approximating kernels converges to the kernel to be approximated. The notion of convergence is given naturally by moving to $\mathbf{BL}_\sigma$ and considering convergence in the Strong Operator Topology (§6).
6. We apply our theory of kernel approximations to two practical applications (§7). First, we show how Bayesian inference can be performed approximately by showing that the †-operation commutes with taking approximations. Secondly, we consider the case of ProbNetKAT, a language developed in [14, 22] to probabilistically reason about networks. ProbNetKAT includes a Kleene star operator $(-)^*$ with a complex semantics which has proved hard to approximate. We show that $(-)^*$ can be approximated, and that the approximation converges.

All the proofs can be found in the Appendix.

Related work. Quasi-Borel sets have recently been proposed as a semantic framework for higher-order probabilistic programs in [24]. The main differences with our approach are: (i) unlike [23, 24] we never leave the realm of kernels, and in particular we never need to worry about normalization. This makes the interpretation of `observe` statements, i.e. of Bayesian inversion, simpler and more natural. However, (ii) unlike the quasi-Borel sets of [24], our category is not Cartesian closed. We can therefore not give a semantics to all higher-order programs. This shortcoming is partly mitigated by the fact that the category of Polish space, on which our category ultimately rests, does have access to many function spaces, in particular all the spaces of functions whose domain is locally compact. We can thus in principle provide a semantics to higher-order programs, provided that λ-abstraction is restricted to locally compact spaces like the reals and the integers, although this won't be investigated in this paper.

The approximation of probabilistic kernels has been a topic of investigation in theoretical computer science for nearly twenty years (see e.g. [4, 9–11]), and for much longer in the mathematical literature (e.g. [5]). Our results build on the formalism developed in [4] with the following differences: (i) we can approximate kernels, their associated *stochastic operator* (backward predicate transformer), or their associated *Markov operator* (forward state transformer) with equivalent ease, and move freely across the three formalisms. (ii) Given a kernel $f : X \to\!\!\!\!\triangleright\, Y$, we can define its approximation $f' : X' \to Y'$ along any quotients X' of X and Y' of Y as in [4], but we can also 'internalize' the approximation as a kernel $f^* : X \to Y$ of the original type. Morally f' and f^* are the same approximation, but the second approximant, being of the same type as the original kernel, can be compared with it. In particular it becomes possible to study the convergence of ever finer approximations, which we do in Section 6. Finally, (iii) we opt to work with Banach lattices rather than the normed cones of [4, 20] because it allows us to formulate the operator side of the theory very naturally, and it connects to a large body of classic mathematical results ([2, 27]) which have been used in the semantics of probabilistic programs as far back as Kozen's seminal [17].

2 A category of Borel kernels

In [6] the first three authors presented a category of Borel kernels similar in spirit to the construction of this section, but with a major shortcoming. As we will shortly see, our category **Krn** of Borel kernels can be equipped with an involutive functor – a *dagger* operation † in the terminology of [21] – which captures the notion of *Bayesian inversion* and is absolutely crucial to everything that follows. In [6] this operation had merely been identified as a map, i.e. not even as a functor. In this section we show that Bayesian inversion does indeed define a †-structure on a more sophisticated – but measure-theoretically very natural – category of kernels.

2.1 Standard Borel spaces and the Giry monad

A *standard Borel space* – or *SB space* for short – is a measurable space $(X, \mathcal{S})$ for which there exists a Polish topology $\mathcal{T}$ on X whose Borel sets are the elements of $\mathcal{S}$, i.e. such that $\mathcal{S} = \sigma(\mathcal{T})$ (see e.g. [16] for an overview). Let us write **SB** for the category of standard Borel spaces and measurable maps. One key structural feature of **SB** is the following:

Theorem 1. *Every* **SB** *object is a limit of a countable co-directed diagram of finite spaces.*

The *Giry monad* was originally defined in two variants [15]:
- As an endofunctor $\mathrm{G}_{\mathbf{Pol}}$ of **Pol**, the category of *Polish spaces*, one sets $\mathrm{G}_{\mathbf{Pol}}(X, \mathcal{T})$ to be the space of Borel probability measures over X together with the weak topology. This space is Polish [16, Th 17.23], and the Portmanteau Theorem [16, Th 17.20]) gives multiple characterizations of the weak topology.
- As an endofunctor $\mathrm{G}_{\mathbf{Meas}}$ of **Meas**, the category of *measurable spaces*, one sets $\mathrm{G}_{\mathbf{Meas}}(X, \mathcal{S})$ to be the set of probability measures on X together with the initial σ-algebra for the maps $ev_A : \mathrm{G}_{\mathbf{Meas}}(X, \mathcal{S}) \to \mathbb{R}, \mu \mapsto \mu(A), A \in \mathcal{S}$.

[2] To the best of our knowledge.

In both cases the Giry monad is defined on an arrow $f : X \to Y$ as the map f_* which sends a measure μ on X to the *pushforward measure* $f_*\mu$ on Y, defined as $G(f)(\mu)(B) = f_*\mu(B) := \mu(f^{-1}(B))$ for B a measurable subset of Y.

We want to define the Giry monad on the category **SB** of standard Borel spaces (and measurable maps), and the two versions of the Giry monad described above offer us natural ways to do this: given an SB space $(X, \sigma(\mathcal{T}))$ we can either compute $\mathrm{G}_{\mathbf{Pol}}(X, \mathcal{T})$ and take the associated standard Borel space, or directly compute $\mathrm{G}_{\mathbf{Meas}}(X, \sigma(\mathcal{T}))$. Fortunately, the two methods agree.

Theorem 2 ([16], Th 17.24)**.** *Let* $\mathrm{B} : \mathbf{Pol} \to \mathbf{SB}$ *denote the functor sending a Polish space* $(X, \mathcal{T})$ *to its associated SB-space* $(X, \sigma(\mathcal{T}))$ *and leaving morphisms unchanged, then*

$$\mathrm{G}_{\mathbf{Meas}} \circ \mathrm{B} = \mathrm{B} \circ \mathrm{G}_{\mathbf{Pol}}.$$

We define the *Giry monad* on SB spaces to be the endofunctor $\mathrm{G} : \mathbf{SB} \to \mathbf{SB}$ defined by either of the two equivalent constructions above. The monadic data of G is given at each SB space X by the unit $\delta_X : X \to \mathrm{G}X, x \mapsto \delta_x$, the Dirac δ measure at x, and the multiplication $m_X : \mathrm{G}^2X \to \mathrm{G}X, \mathbb{P} \mapsto \lambda A. \int_{\mathrm{G}X} ev_A d\mathbb{P}$. We refer the reader to [15] for proofs that δ_X and m_X are measurable.

2.2 The construction of Krn

Let us denote by $\mathbf{SB}_{\mathrm{G}}$ the Kleisli category associated with the Giry monad (G, δ, m). We denote Kleisli arrows, i.e. Markov kernels, by $X \rightarrowtriangle Y$, and we call such an arrow *deterministic* if it can be factorized as an ordinary measurable function followed by the unit δ. Kleisli composition is denoted by $\bullet$. The category $* \downarrow \mathbf{SB}_{\mathrm{G}}$ has arrows $* \rightarrowtriangle X$ as objects, where $*$ is the one point SB space (the terminal object in **SB**). An arrow from $\mu : * \rightarrowtriangle X$ to $\nu : * \rightarrowtriangle Y$ is a $\mathbf{SB}_{\mathrm{G}}$ arrow $f : X \rightarrowtriangle Y$ such that $\nu = f \bullet \mu$, i.e. such that $\nu(A) = \int_X f(x)(A) d\mu$ for any measurable subset A of Y. This situation will be denoted in short by $f : (X, \mu) \rightarrowtriangle (Y, \nu)$, and we will call a pair (X, μ) a *measured SB space.*

We want to construct a quotient of $* \downarrow \mathbf{SB}_{\mathrm{G}}$, such that two $* \downarrow \mathbf{SB}_{\mathrm{G}}$ arrows are identified if they disagree on a null set w.r.t. the measure on their domain. For $g, g' : (X, \mu) \rightarrowtriangle (Y, \nu)$, we define $N(g, g') = \{x \in X \mid g(x) \neq g'(x)\}$.

Lemma 3. *$N(g, g')$ is a measurable set.*

We now define a relation $\sim$ on $\mathrm{Hom}((X, \mu), (Y, \nu))$ by saying that for any two arrows $g, g' : (X, \mu) \rightarrowtriangle (Y, \nu)$, $g \sim g'$ if $\mu(N(g, g')) = 0$. This clearly defines an equivalence relation on $\mathrm{Hom}((X, \mu), (Y, \nu))$. In order to perform the quotient of the category $* \downarrow \mathbf{SB}_{\mathrm{G}}$ modulo $\sim$, we need to check that it is compatible with composition.

Proposition 4. *If $g \sim g'$, then $h \bullet g \bullet f \sim h \bullet g' \bullet f$.*

Definition 5. *Let* **Krn** *be the category obtained by quotienting $* \downarrow \mathbf{SB}_{\mathrm{G}}$ hom-sets with $\sim$.*

The following Theorem is of great practical use and generalizes the well-known result for deterministic arrows.

Theorem 6 (Change of Variables in **Krn**)**.** *Let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ be a* **Krn***-morphism. For any measurable function $\phi : Y \to \mathbb{R}$, if ϕ is ν-integrable, then $\phi \bullet f(x) = \int_Y \phi \, df(x)$ is μ-integrable and*

$$\int_Y \phi \, d\nu = \int_X \phi \bullet f \, d\mu$$

The symmetric monoidal structure of **Krn** is defined on a pair of objects $(X, \mu), (Y, \nu)$ by the Cartesian product and the product of measure, i.e. $(X, \mu) \otimes (Y, \nu) = (X \times Y, \mu \otimes \nu)$. On pairs of morphisms $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ and $f' : (X', \mu') \rightarrowtriangle (Y', \nu')$ it is defined by $(f \otimes f')(x, x') := f(x) \otimes f'(x')$. The unitors, associator and braiding transformations are given by the obvious bijections.

2.3 The dagger structure of Krn

Krn has an extremely powerful inversion principle:

Theorem 7 (Measure Disintegration Theorem, [16], 17.35)**.** *Let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ be a* deterministic **Krn***-morphism, there exists a unique morphism $f^\dagger_\mu : (Y, \nu) \rightarrowtriangle (X, \mu)$ such that*

$$f \bullet f^\dagger_\mu = \mathrm{id}_{(Y,\nu)}. \tag{1}$$

The kernel $f^\dagger_\mu$ is called the *disintegration of μ along f*. As our notation suggests, the disintegration depends fundamentally on the measure μ over the domain, however we will omit this subscript when there is no ambiguity. The following lemma relates disintegrations to conditional expectations.

Lemma 8 ([8])**.** *Let $f : (X, \mu) \to (Y, \nu)$ be a deterministic* **Krn***-morphism, and let $\phi : X \to \mathbb{R}$ be measurable, then μ-a.e.*

$$\phi \bullet f^\dagger \bullet f = \mathbb{E}[\phi \mid \sigma(f)]$$

We can extend the definition of $(-)^\dagger$ to *any* **Krn**-morphism $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ in a functorial way, although $f^\dagger$ will not in general be a right inverse to f. The construction of $f^\dagger$ is detailed in [6], but let us briefly recall how it works. The category **SB** has products which are built in the same way as in **Meas** via the product of σ-algebras[3]. Given any kernel $f : (X, \mu) \rightarrowtriangle (Y, \nu)$, we can canonically construct a probability measure γ_f on the product $X \times Y$ of SB-space by defining it on the rectangles of $X \times Y$ as

$$\gamma_f(A \times B) = \int_{x \in X} 1_A(x) \cdot f(x)(B) \, d\mu. \tag{2}$$

Equivalently, $\gamma_f = (\delta_X \otimes f) \bullet \Delta_X \bullet \mu$, where $\Delta_X : X \to X \times X$ is the diagonal map. Letting $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ be the canonical projections, we observe that $\mathrm{G}\pi_X(\gamma_f) = \mu$ and $\mathrm{G}\pi_Y(\gamma_f) = \nu$: in other words, γ_f is a coupling of μ and ν. The disintegration of γ_f along π_Y is a kernel $\pi^\dagger_Y : (Y, \nu) \to (X \times Y, \gamma_f)$. Finally we define:

$$f^\dagger = \pi_X \bullet \pi^\dagger_Y. \tag{3}$$

[3]Unlike the category **Krn** which does *not* have products.

The following **Krn** diagram sums up the situation:

$$\begin{tikzcd}
(X,\mu) \arrow[r, bend left, "\pi_X^\dagger"] \arrow[rr, bend left=60, "f"] & (X\times Y, \gamma_f) \arrow[l, bend left, "\pi_X"] \arrow[r, bend left, "\pi_Y"] & (Y,\nu) \arrow[l, bend left, "\pi_Y^\dagger"] \arrow[ll, bend left=60, "f^\dagger"]
\end{tikzcd}$$

where $\pi_X^\dagger$ is explicitly given by $(\delta_X \otimes f) \bullet \Delta_X$. The following property characterizes the action of $(-)^\dagger$ on **Krn**-morphisms:

Theorem 9. *For all $f : (X,\mu) \rightarrowtriangle (Y,\nu)$, $f^\dagger : (Y,\nu) \rightarrowtriangle (X,\mu)$ is the unique* **Krn** *morphism satisfying for all measurable sets $A \subseteq X$, $B \subseteq Y$ the following equation:*

$$\int_{x\in X} 1_A(x)\cdot f(x)(B)\, d\mu = \int_{y\in Y} f^\dagger(y)(A)\cdot 1_B(y)\, d\nu \quad (4)$$

In view of Eq. (4), we will call $f^\dagger$ the *Bayesian inversion of f*, and refer to $(-)^\dagger$ as the *Bayesian inversion operation* on **Krn**. It will be crucial throughout the rest of this paper. It is important to see that $f^\dagger$ absolutely depends on the choice of μ and not only on f seen as a function. We can now improve on [6] and show that $(-)^\dagger$ is indeed a †-operation in the strict categorical meaning of the term.

Theorem 10. **Krn** *is a dagger symmetric monoidal category, with $(-)^\dagger$ given by Bayesian inversion.*

3 Banach lattices

It is well-known that kernels can alternatively be seen as predicate – i.e. real-valued function –transformers, or as state – i.e. probability measure – transformers. The latter perspective was adopted by Kozen in [17] to describe the denotational semantics of probabilistic programs (without conditioning). We shall see in this section and the next, that the predicate and state transformer perspectives are dual to one another in the category of *Banach lattices*, a framework incidentally also used in [17]. For an introduction to the theory of Banach lattices we refer the reader to e.g. [2, 27].

An *ordered real vector space* V is a real vector space together with a partial order $\leq$ which is compatible with the linear structure in the sense that for all $u, v, w \in V, \lambda \in \mathbb{R}^+$

$$u \leq v \Rightarrow u + w \leq v + w \qquad \text{and} \qquad u \leq v \Rightarrow \lambda u \leq \lambda v$$

An ordered vector space $(V, \leq)$ is called a *Riesz space* if the poset structure forms a lattice. A vector v in a Riesz space $(V, \leq)$ is called *positive* if $0 \leq v$, and its *absolute value* $|v|$ is defined as $|v| = v \vee (-v)$. A Riesz space $(V, \leq)$ is *σ-order complete* if every non-empty countable subset of V which is order bounded has a supremum.

A *normed Riesz space* is a Riesz space $(V, \leq)$ equipped with a *lattice norm*, i.e. a map $\|\cdot\| : V \to \mathbb{R}$ such that:

$$|v| \leq |w| \text{ implies } \|v\| \leq \|w\|\,. \quad (5)$$

A normed Riesz space is called a *Banach lattice* if it is (norm-) complete, i.e. if every Cauchy sequence (for the norm $\|\cdot\|$) has a limit in V.

Example 11. *For each measured space (X,μ) – and in particular* **Krn***-objects – and each $1 \leq p \leq \infty$, the space $\mathrm{L}_p(X,\mu)$ is a Riesz space with the pointwise order. When it is equipped with the usual L_p-norm, it is a Banach lattice. This fact is often referred to as the* Riesz-Fischer theorem *(see [2, Th 13.5]). We will say that $p, q \in \mathbb{N} \cup \{\infty\}$ are* Hölder conjugate *if either of the following conditions hold: (i) $1 < p, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, or (ii) $p = 1$ and $q = \infty$, or (iii) $p = \infty$ and $q = 1$.*

Theorem 12 (Lemma 16.1 and Theorem 16.2 of [27])**.** *Every Banach lattice is σ-order complete.*

There are two very natural modes of 'convergence' in a Banach lattice: *order convergence* and *norm convergence*. The latter is well-known, the former less so. An order bounded sequence $\{v_n\}_{n\in\mathbb{N}}$ in a σ-complete Riesz space (and thus in a Banach lattice) *converges in order to v* if either of the following equivalent conditions holds:

$$v = \liminf_n v_n := \bigvee_n \bigwedge_{n\leq m} v_m, \quad v = \limsup_n v_n := \bigwedge_n \bigvee_{n\leq m} v_m.$$

For a monotone increasing sequence v_n, this definition simplifies to $v = \bigvee_n v_n$, which is often written $v_n \uparrow v$.

In a general σ-complete Riesz space, order and norm convergence are disjoint concepts, i.e. neither implies the other (see [27, Ex. 15.2] for two counter-examples). However if a sequence converges both in order and in norm then the limits are the same (see [27, Th. 15.4]). Moreover, for *monotone* sequences norm convergence implies order convergence:

Proposition 13 ([27] Theorem 15.3)**.** *If $\{v_n\}_{n\in\mathbb{N}}$ is an increasing sequence in a normed Riesz space and if v_n converges to v in norm (notation $v_n \to v$), then $v_n \uparrow v$.*

In a Banach lattice we have the following stronger property.

Proposition 14 (Lemma 16.1 and Theorem 16.2 of [27])**.** *If $\{v_n\}_{n\in\mathbb{N}}$ is a sequence of positive vectors in a Banach lattice such that $\sup_n \|v_n\|$ converges, then $\bigvee_n v_n$ exists and $\|\bigvee_n v_n\| = \bigvee_n \|v_n\|$.*

It can also happen that order convergence implies norm convergence. A lattice norm on a Riesz space is called *σ-order continuous* if $v_n \downarrow 0$ (v_n is a decreasing sequence whose infimum is 0) implies $\|v_n\| \downarrow 0$.

Example 15. *For $1 \leq p < \infty$, the L_p-norm is σ-*order continuous*, and thus order convergence and norm convergence coincide. However, for $p = \infty$ this is not the case as the following simple example shows. Consider the sequence of essentially bounded functions $v_n = 1_{[n,+\infty[}$: it is decreasing for the order on $\mathrm{L}_\infty(\mathbb{R},\lambda)$ with the constant function 0 as its infimum, i.e. $v_n \downarrow 0$. However $\|v_n\| = 1$ for all n.*

Many types of morphisms between Banach lattices are considered in the literature but most are at least *linear* and *positive*, that is to say they send positive vectors to positive vectors. From now on, we will assume that all morphisms

are positive (linear) operators. Other than that, we will only mention two additional properties, corresponding to the two modes of convergence which we have examined. The first notion is very well-known: a linear operator $T : V \to W$ between normed vector spaces is called *norm-bounded* if there exists $C \in \mathbb{R}$ such that $\|Tv\| \leq C\,\|v\|$ for every $v \in V$. The following result is familiar:

Theorem 16. *An operator $T : V \to W$ between normed vector spaces is norm-bounded iff it is continuous.*

Thus norm-bounded operators preserve norm-convergence. The corresponding order-convergence concept is defined as follows: an operator $T : V \to W$ between σ-order complete Riesz spaces is said to be *σ-order continuous* if whenever $v_n \uparrow v$, $Tv = \bigvee Tv_n$. It follows that we can consider two types of dual spaces on a Banach lattice V: on the one hand we can consider the *norm-dual*:

$$V^* = \{f : V \to \mathbb{R} \mid f \text{ is norm-continuous}\}$$

and the *σ-order-dual*:

$$V^\sigma = \{f : V \to \mathbb{R} \mid f \text{ is } \sigma\text{-order continuous}\}$$

The latter is sometimes known as the *Köthe dual* of V (see [12, 27]). The two types of duals coincide for a large class of Banach spaces of interest to us.

Theorem 17. *If a Banach lattice V admits a strictly positive linear functional and has a σ-order-continuous norm, then $V^* = V^\sigma$.*

Example 18. *The result above can directly be applied to our running example: given a measured space (X, μ) and an integer $1 \leq p < \infty$, the Lebesgue integral provides a strictly positive functional on $\mathrm{L}_p(X, \mu)$, and we already know from Example 15 that $\mathrm{L}_p(X, \mu)$ has a σ-order-continuous norm. It follows that*

$$\mathrm{L}_p(X, \mu)^* = \mathrm{L}_p(X, \mu)^\sigma$$

Moreover, it is well-known that if (p, q) are Hölder conjugate and $1 < p, q < \infty$, then $\mathrm{L}_p(X, \mu)^* = \mathrm{L}_q(X, \mu)$, and thus $\mathrm{L}_p(X, \mu)^\sigma = \mathrm{L}_q(X, \mu)$. It is also known that $\mathrm{L}_1(X, \mu)^* = \mathrm{L}_\infty(X, \mu)$, and thus $\mathrm{L}_1(X, \mu)^\sigma = \mathrm{L}_\infty(X, \mu)$.

However Theorem 17 does not *hold for $\mathrm{L}_\infty(X, \mu)$ since the L_∞-norm is not σ-order continuous, as was shown in Example 15. It is well-known that $\mathrm{L}_\infty(X, \mu)^* \neq \mathrm{L}_1(X, \mu)$, and in fact $\mathrm{L}_\infty(X, \mu)^*$ can be concretely described as the Banach lattice* $\mathrm{ba}(X, \mu)$ *of* charges *(i.e. finitely additive finite signed measures) which are absolutely continuous w.r.t, μ on X (see [13, IV.8.16]). However, as is shown in e.g. [4, 27]*

$$\mathrm{L}_\infty(X, \mu)^\sigma = \mathrm{L}_1(X, \mu) \tag{6}$$

As Examples 15 and 18 show, the $(-)^\sigma$ operation brings a lot of symmetry to the relationship between L_p-spaces since

$$\mathrm{L}_p(X, \mu)^\sigma = \mathrm{L}_q(X, \mu)$$

for *any* Hölder conjugate pair $1 \leq p \leq \infty$. For this reason we will consider the category $\mathbf{BL}_\sigma$ whose objects are Banach lattices and whose morphisms are σ-order continuous positive operators. Note that the Köthe dual of a Banach lattice is a Banach lattice, and it easily follows that $(-)^\sigma$ in fact defines a contravariant functor $\mathbf{BL}_\sigma^{\mathrm{op}} \to \mathbf{BL}_\sigma$ which acts on morphisms by pre-composition. As we will now see, $\mathbf{BL}_\sigma$ is the category in which predicate and state transformers are most naturally defined.

4 From Borel kernels to Banach lattices

The functors S_p and T_p. For $1 \leq p \leq \infty$, the operation which associates to a **Krn**-object (X, μ) the space $\mathrm{L}_p(X, \mu)$ can be thought of as either a contravariant or a covariant functor. We define the functors $\mathsf{S}_p : \mathbf{Krn} \to \mathbf{BL}_\sigma^{\mathrm{op}}, 1 \leq p \leq \infty$ as expected on objects, and on **Krn**-morphisms $f : X \rightarrowtriangle Y$ via the well-known 'predicate transformer' perspective:

$$\mathsf{S}_p(f) : \mathrm{L}_p(Y, \nu) \to \mathrm{L}_p(X, \mu), \phi \mapsto \lambda x. \int_Y \phi \, df(x) = \phi \bullet f$$

For a proof that this defines a functor see [6]. We define the covariant functors $\mathsf{T}_p : \mathbf{Krn} \to \mathbf{BL}_\sigma, 1 \leq p \leq \infty$ as $\mathsf{T}_p = \mathsf{S}_p \circ (-)^\dagger$.

The functor $\mathcal{M}^{\ll\cdot}$. An *ideal* of a Riesz space V is a subvector space $U \subseteq V$ with the property that if $|u| \leq |v|$ and $v \in U$ then $u \in U$. An ideal U is called a *band* when for every subset $D \subseteq U$ if $\bigvee D$ exists in V, then it also belongs to U. Every band in a Banach lattice is itself a Banach lattice. Of particular importance is the band B_v generated by a singleton $\{v\}$, which can be described explicitly as

$$B_v = \{w \in V \mid (|w| \wedge n\,|v|) \uparrow |w|\}$$

Example 19. *Let X be an SB-space and* $\mathrm{ca}(X)$ *denote the set of measures of bounded variation on X. It can be shown ([2, Th 10.56]) that* $\mathrm{ca}(X)$ *is a Banach lattice. The linear structure on* $\mathrm{ca}(X)$ *is as expected, the Riesz space structure is given by*

$$(\mu \vee \nu)(A) = \sup\{\mu(B) + \nu(A \setminus B) \mid B \text{ measurable}, B \subseteq A\}$$

and the dual definition for the meet operation. The norm is given by the total variation i.e.

$$\|\mu\| = \sup\left\{\sum_i^n |\mu(A_i)| \,\middle|\, \{A_1, \ldots, A_n\} \text{ a meas. partition of } X\right\}$$

Given $\mu \in \mathrm{ca}(X)$, the band B_μ generated by μ is just the set of measures of bounded variation which are absolutely continuous w.r.t. μ. In particular B_μ is a Banach lattice.

We can now define the functor $\mathcal{M}^{\ll\cdot} : \mathbf{Krn} \to \mathbf{BL}_\sigma$ by:

$$\begin{cases} \mathcal{M}^{\ll\cdot}(X, \mu) := B_\mu \\ \mathcal{M}^{\ll\cdot} f : \mathcal{M}^{\ll\cdot}(X, \mu) \to \mathcal{M}^{\ll\cdot}(Y, \nu), \rho \mapsto f \bullet \rho \end{cases}$$

We will usually write $\mathcal{M}^{\ll\cdot}(X, \mu)$ as $\mathcal{M}^{\ll\mu}(X)$.

Proposition 20. *Let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ be a* **Krn** *arrow. Let ρ be a finite measure on X such that $\rho \ll \mu$. Then $f \bullet \rho \ll \nu$, and thus $\mathcal{M}^{\ll\cdot}$ defines a functor.*

Radon-Nikodym is natural. We now present a first pair of natural transformations which will establish a natural isomorphism between the functors T_1 and $\mathcal{M}^{\ll\cdot}$. First, we

define the *Radon-Nikodym transformation* $\mathsf{rn} : \mathcal{M}^{\ll\cdot} \to \mathsf{T}_1$ at each **Krn**-object (X, μ) by the map

$$\mathsf{rn}_{(X,\mu)} : \mathcal{M}^{\ll\mu}(X) \to \mathrm{L}_1(X, \mu), \quad \mathsf{rn}_{(X,\mu)}(\rho) = \frac{d\rho}{d\mu}$$

where $d\rho/d\mu$ is of course the Radon-Nikodym derivative of ρ w.r.t. μ. The fact that this transformation defines a positive operator between Banach lattices is simply a restatement of the usual Radon-Nikodym theorem [13, III.10.7.], combined with the well-known linearity property of the Radon-Nikodym derivative. To see that it is also σ-order-continuous, consider a monotone sequence $\mu_n \uparrow \mu$ converging in order to μ in $\mathcal{M}^{\ll\nu}(X)$. This means that for any measurable set A of X, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$. Since $(d\mu_n/d\nu)_{n\in\mathbb{N}}$ is bounded in L_1-norm the function $g = \bigvee_n d\mu_n/d\nu$ exists and is simply the pointwise limit $g(x) = \lim_{n\to\infty} d\mu_n/d\nu(x)$. It now follows from the monotone convergence theorem (MCT) that

$$\int_A g d\nu = \int_A \lim_{n\to\infty} \frac{d\mu_n}{d\nu} d\nu = \lim_{n\to\infty} \int_A \frac{d\mu_n}{d\nu} d\nu = \lim_{n\to\infty} \mu_n(A) = \mu(A)$$

in other words, $g = d\mu/d\nu$ and rn is well-defined. That rn is also natural has – to our knowledge – never been published.

Theorem 21. *The Radon-Nikodym transformation is natural.*

Secondly, we define the *Measure Representation transformation* $\mathsf{mr} : \mathsf{T}_1 \to \mathcal{M}^{\ll\cdot}$ at each **Krn**-object (X, μ) by the map $\mathsf{mr}_{(X,\mu)} : \mathsf{T}_1(X, \mu) \to \mathcal{M}^{\ll\cdot}(X, \mu)$ defined as

$$\mathsf{mr}_{(X,\mu)}(f)(B_X) = \int_{B_X} f d\mu$$

This is a very well-known construction in measure theory, and the fact that $\mathsf{mr}_{(X,\mu)}$ is a σ-order continuous operator between Banach lattices is immediate from the linearity of integrals and the MCT.

Theorem 22. *The Measure Representation transformation is natural.*

Riesz representations are natural. We now present a second pair of natural transformations which will establish a natural isomorphism between $(-)^\sigma \circ \mathrm{S}_\infty$ and $\mathcal{M}^{\ll\cdot}$. First, we define the *Riesz Representation* transformation $\mathsf{rr} : (-)^\sigma \circ \mathrm{S}_\infty \to \mathcal{M}^{\ll\cdot}$ at each **Krn**-object (X, μ) by the map $\mathsf{rr}_{(X,\mu)} : (-)^\sigma \circ \mathrm{S}_\infty(X, \mu) \to \mathcal{M}^{\ll\cdot}(X, \mu)$ defined as

$$\mathsf{rr}_{(X,\mu)}(F)(B_X) = F(1_{B_X})$$

This construction is key to a whole collection of results in functional analysis commonly known as Riesz Representation Theorems (see [2] Chapter 14 for an overview). One can readily check that the Riesz Representation transformation is well-defined: $\mathsf{rr}_{(X,\mu)}(F)(\emptyset) = F(0) = 0$ and the σ-additivity of $\mathsf{rr}_{(X,\mu)}(F)$ follows from the σ-order-continuity of F. To see that $\mathsf{rr}_{(X,\mu)}(F) \ll \mu$, assume that $\mu(B_X) = 0$, then clearly $1_{B_X} = 0$ μ-a.e., i.e. $1_{B_X} = 0$ in $\mathrm{L}_\infty(X, \mu)$, and thus $F(1_{B_X}) = 0$.

Theorem 23. *The Riesz Representation transformation is natural.*

Finally, we define the *Functional Representation transformation* fr at each **Krn**-object (X, μ) by the map $\mathsf{fr}_{(X,\mu)} : \mathcal{M}^{\ll\cdot}(X, \mu) \to (-)^\sigma \circ \mathrm{S}_\infty(X, \mu)$ by

$$\mathsf{fr}_{(X,\mu)}(\mu)(\phi) = \int_X \phi d\mu$$

This construction is also completely standard in measure theory, although it has never to our knowledge been seen as a natural transformation.

Theorem 24. *The Functional Representation transformation is well-defined, i.e.* $\mathsf{fr}_{(X,\mu)}$ *is a σ-order continuous positive operator, and is natural.*

Natural Isomorphisms We have now defined the following four natural transformations:

$$\mathsf{T}_1 \underset{\mathsf{rn}}{\overset{\mathsf{mr}}{\rightleftarrows}} \mathcal{M}^{\ll\cdot} \underset{\mathsf{rr}}{\overset{\mathsf{fr}}{\rightleftarrows}} (\mathrm{S}_\infty)^\sigma$$

In fact, both pairs form natural isomorphisms, and these can be restricted to arbitrary Hölder conjugate pairs (p, q).

Theorem 25. rn *and* mr *are inverse of one another, in particular there exists a* natural *isomorphism between* $\mathcal{M}^{\ll\mu}(X)$ *and* $\mathrm{L}_1(X, \mu)$.

Theorem 26. rr *and* fr *are inverse of each other, in particular there exists a* natural *isomorphism between* $\mathcal{M}^{\ll\mu}(X)$ *and* $(\mathrm{L}_\infty(X, \mu))^\sigma$.

We can now conclude that the isomorphism proved in Theorem 6 of [6] is in fact natural.

Corollary 27. *There exists a* natural *isomorphism between* $\mathsf{T}_1 := \mathrm{S}_1 \circ (-)^\dagger$ *and* $(-)^\sigma \circ \mathrm{S}_\infty$.

We can in fact restrict this result to *any* Hölder conjugate pair (p, q):

Theorem 28. *For* $1 \le p \le \infty$ *with Hölder conjugate* q*, the natural transformation* $\mathsf{rn} \circ \mathsf{rr}$ *restricts to a natural transformation* $(-)^\sigma \circ \mathrm{S}_q \to \mathsf{T}_p$.

The correspondence between the various categories and functors discussed in this section are summarized as follows:

$$\begin{tikzcd}
 & \mathbf{BL}_\sigma & \\
\mathbf{BL}_\sigma^{\mathrm{op}} \arrow[ur, "(-)^\sigma"] \arrow[rr, shift right, "\mathsf{rr}"'] & & \mathbf{Krn}^{\mathrm{op}} \arrow[ul, "\mathrm{S}_1"'] \arrow[ll, shift right, "\mathsf{fr}"'] \arrow[ll, shift left, "\mathsf{mr}"] \arrow[ll, shift right, "\mathsf{rn}"'] \arrow[rr, shift left, "\mathsf{rn}"] \\
 & \mathbf{Krn} \arrow[ul, "\mathrm{S}_\infty"] \arrow[ur, "(-)^\dagger"'] \arrow[uu, "\mathcal{M}^{\ll\cdot}"] &
\end{tikzcd} \tag{7}$$

5 Approximations

In this section we develop a scheme for approximating kernels which follows naturally from the †-structure of **Krn**. Consider $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ and a pair of deterministic maps $p : (X, \mu) \to (X', p_*\mu)$ and $q : (Y, \nu) \to (Y', q_*\nu)$ (typically these maps coarsen the spaces X and Y).

$$\begin{tikzcd} (X,\mu) \arrow[r, "f"] \arrow[r, shift right, "f^{p,q}"'] \arrow[d, bend right, "p"'] & (Y,\nu) \arrow[d, bend right, "q"'] \\ (X',p_*\mu) \arrow[r, "f_{p,q}"'] \arrow[u, bend right, "p^\dagger_\mu"'] & (Y',q_*\nu) \arrow[u, bend right, "q^\dagger_\nu"'] \end{tikzcd} \tag{8}$$

The †-structure of **Krn** allows us to define the new kernels

$$f^{p,q} := q^\dagger_\nu \bullet q \bullet f \bullet p^\dagger_\mu \bullet p \qquad : X \rightarrowtriangle Y \tag{9}$$

$$f_{p,q} := q \bullet f \bullet p^\dagger_\mu \qquad : X' \rightarrowtriangle Y' \tag{10}$$

The supscript notation is meant to indicate that the approximation lives 'upstairs' in Diagram (8) and conversely for the subscripts. Intuitively, $f_{p,q}$ and $f^{p,q}$ take the average of f over the fibres given by p, q according to μ and ν (see Section 7 for concrete calculations). The advantage of (10) is that we can approximate a kernel on a huge space by a kernel on a, say, finite one. The advantage of (9) is that although it is more complicated, it is morally equivalent and has the same type as f, which means that we can compare it to f.

A very simple consequence of our definition is that Bayesian inversion commutes with approximations. We shall use this in §7.1 to perform *approximate Bayesian inference.*

Theorem 29. *Let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$, let $p : X \to X'$ and $q : Y \to Y'n$ be a pair of deterministic maps, then*

$$(f^\dagger)^{q,p} = (f^{p,q})^\dagger \text{ and } (f^\dagger)_{q,p} = (f_{p,q})^\dagger$$

In practice we will often consider endo-kernels $f : X \rightarrowtriangle X$ with a single coarsening map $p : X \to X'$ to a *finite space.* In this case (9) simplifies greatly.

Proposition 30. *Under the situation described above*

$$f^p := p^\dagger_\nu \bullet p \bullet f \bullet p^\dagger_\mu \bullet p = f \bullet p^\dagger_\mu \bullet p \tag{11}$$

In the case covered by Proposition 30, the interpretation of f^p is very natural: for each $x \in X$ the measure $f(x)$ is approximated by its average over the fibre to which x belongs, conditioned on being in the fibre. For fibres with strictly positive μ-probability, this is simply

$$f^p(x)(A) = \frac{\int_{y \in p^{-1}(p(x))} f(y)(A)\, d\mu}{\mu(p^{-1}(p(x))}$$

However (11) also covers the case of μ-null fibres. Note also that in the case where $f^p = f$, the map p corresponds to what is known as a strong functional bisimulation for f.

Approximating is non-expansive. It is well-known that conditional expectations are non-expansive and we know from Lemma 8 that pre-composing by $p^\dagger_\mu \bullet p$ as in (11) amounts to conditioning. The following lemma is an easy consequence.

Lemma 31. *Let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ and $q : X \to X'$ be a deterministic quotient, then for all $1 \le p \le \infty$ and $\phi \in \mathrm{L}_p(Y, \nu)$*

$$\left\|\mathrm{S}_p f^q(\phi)\right\|_p \le \left\|\mathrm{S}_p f(\phi)\right\|_p$$

Compositionality of approximations In the case where we wish to approximate a composite kernel $g \bullet f$, it might be convenient, for modularity reasons, to approximate f and g separately. This does not entail any loss of information provided the quotient maps are hemi-bisimulations, in the following sense. Let $p : X \to X', q : Y \to Y', r : Z \to Z'$ be deterministic quotients and let $f : (X, \mu) \rightarrowtriangle (Y, \nu), g : (Y, \nu) \rightarrowtriangle (Z, \rho)$ be composable kernels. We say that q is a left hemi-bisimulation for f if $f = q^\dagger \bullet q \bullet f$, and conversely that it is a right hemi-bisimulation for g if $g = g \bullet q^\dagger \bullet q$ holds. In either case, one can verify using Theorems 7 and 29 that approximation commutes with composition, i.e. that $(g \bullet f)^{p,r} = g^{q,r} \bullet f^{p,q}$.

Discretization schemes We will use (10) and (11) to build sequences of arbitrarily good approximations of kernels. For this we introduce the following terminology.

Definition 32. *We define a* discretization scheme *for an SB-space X to be a countable co-directed diagram (ccd) of finite spaces for which X is a cone (not necessarily a limit).*

If $(X_i)_{i \in I}$ is a discretization scheme of X and $p_i : X \to X_i$ are the maps making X a cone, then it follows from the definition that if $i < j$, $\sigma(p_i) \subseteq \sigma(p_j)$ where $\sigma(p_i)$ is the σ-algebra generated by p_i. For each $i \in I$ the finite quotient p_i defines a *measurable partition* of X whose disjoint components $p_i^{-1}(\{k\}), k \in X_i$ we will call *cells.*

By Theorem 1 every SB-space has a discretization scheme for which it is not just a cone but a limit.

In practice we will work with discretization schemes linearly ordered by $\mathbb{N}$. In this case the sequence $(X, \sigma(p_n))_{n \in \mathbb{N}}$ defines what probabilists call a *filtration* and we will denote the approximation f^{p_n} given by (11) simply by f^n.

6 Convergence

We now turn to the question of convergence of approximations. There appears to be little literature on the subject of the convergence of approximations of Markov kernels. One rare reference is [5]. Via the functor S_p defined above in Sections 3 and 4 we can seek a topology in terms of the *operators associated to a sequence of kernels.* Indeed, following [5], we will prove convergence results for the *Strong Operator Topology* (SOT).

Definition 33. *We will say that a sequence of kernels $f^n : X \rightarrowtriangle Y$* converges to *$f : X \rightarrowtriangle Y$* in strong operator topology, *and write $f^n \longrightarrow_{\mathrm{s}} f$, if $\mathrm{S}_1 f^n$ converges to $\mathrm{S}_1 f$ in the strong*

operator topology, i.e. if

$$\lim_{n\to\infty} \|\mathrm{S}_1 f^n(\phi) - \mathrm{S}_1 f(\phi)\|_1 = 0$$

Proving convergence. We start with the following key lemma which is a consequence of Lévy's upward convergence Theorem ([25, Th. 14.2]) .

Lemma 34. *Let $f : (X, \mu) \rightarrowtail (Y, \nu)$ be a* **Krn***-morphism and let $p_n : X \to X_n, n \in \mathbb{N}$ be a discretization scheme such that for $\mathcal{B}_X$ the Borel σ-algebra of X we have*

$$\mathcal{B}_X = \sigma\left(\bigcup_n \sigma(p_n)\right)$$

and let $A \subseteq Y$ be measurable, then for $f^n := f \bullet p_n^\dagger \bullet p_n$

$$\lim_{n\to\infty} f^n(x)(A) = f(x)(A)$$

for μ-almost every $x \in X$. Moreover,

$$\lim_{n\to\infty} \|\mathrm{S}_1 f^n(1_A) - \mathrm{S}_1 f(1_A)\|_1 = 0$$

Theorem 35 (Convergence of Approximations Theorem)**.** *Under the conditions of Lemma 34, for μ-almost every $x \in X$*

$$\lim_{n\to\infty} f^n(x)(A) = f(x)(A)$$

for all Borel subsets A. Moreover,

$$\lim_{n\to\infty} \|\mathrm{S}_1 f^n(\phi) - \mathrm{S}_1 f(\phi)\|_1 = 0$$

for any $\phi \in \mathrm{L}_1(X, \nu)$. In other words $f^n \longrightarrow_s f$.

Note that operators of the shape $\mathrm{S}_p f^n$ obtained from a discretization scheme are *finite rank operators*. Thus, we, in fact, also obtained a theorem to approximate stochastic operators by *stochastic operators of finite rank* for the SOT topology. In general, we cannot hope for convergence in the stronger norm topology since the identity operator – which is stochastic – is a limit of operators of finite rank in the norm topology iff the space is finite dimensional.

Note also that the various relationships established in Section 4 allow us to move from an approximation of a kernel to an approximation of the corresponding Markov operator. Since a discretization scheme making $f^n \longrightarrow_s f$ will also make $(f^\dagger)^n \longrightarrow_s f^\dagger$, it follows from Theorem 25 that we get a finite rank approximation of the Markov operator $\mathcal{M}^{\ll\cdot}(f)$.

7 Applications

7.1 Approximate Bayesian Inference

Consider again the inference problem from the introduction. There one needed to invert $f(x) = \mathcal{N}(x, 1)$ with prior $\mu = \mathcal{N}(0, 1)$. We can use Theorem 29 to see how our approximate Bayesian inverse compares to the exact solution which in this simple case is known to be $f_\mu^\dagger(0.5) = \mathcal{N}(1/4, 1/2)$. To do this, we use a doubly indexed discretization scheme:

$$q_{mn} : \mathbb{R} \to 2 \times m \times n + 2$$

defining a window of width $2m$ centred at 0 divided in $2mn$ equal intervals; with the remaining intervals $(-\infty, -m]$ and (m, ∞) each sent to a point (hence the +2 above).

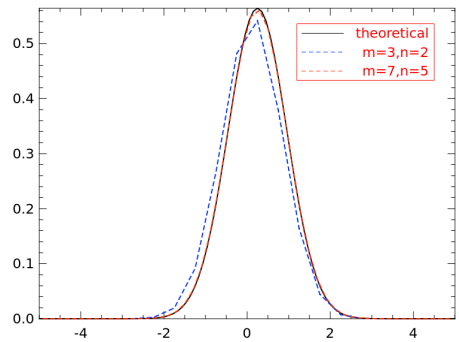

Figure 1. Approximate posteriors

Since all classes induced by q_{mn} have positive μ-mass, approximants can be computed simply as:

$$f^{m,n}([k])([l]) = \mu[k]^{-1} \int_{x\in[k]} \mathcal{N}(x, 1)([l])\, d\mu$$

where $[k]$, $[l]$ range over classes of q_{mn}. The corresponding stochastic matrices are shown in Fig. 2 and 3 for $m, n = 5, 3$ and $6, 10$ respectively.

Since these approximants are finite, their Bayesian inverse can be computed directly by Bayes theorem (i.e. taking the adjoint of the stochastic matrices):

$$f^{m,n\dagger}([l])([k]) = \frac{\mu[k] \cdot f^{m,n}([k])([l])}{\nu[l]} \tag{12}$$

with $\nu = f_*(\mu)$. Commutation of inversion and approximation guarantees that the $f^{m,n\dagger}$ converge to $f^\dagger$.

Indeed, Fig. 1 shows the the Lebesgue density of $f^{m,n\dagger}(0.5)$ for $m, n = 3, 2$ (in dashed blue) and $7, 5$ (dashed red). The latter approximant is already hardly distinguishable from the exact solution (solid black).

It must be emphasized that this example is meant only as an illustration and does not constitute a universal solution to the irreducibly hard (not even computable in general [1]) problem of performing Bayesian inversion. Also, not all quotients are equally convenient: what makes the approach computationally tractable is that the fibres are easily described and the measure conveniently evaluated on such fibres.

7.2 Approximating the Kleene star of ProbNetKAT

ProbNetKAT ([14, 22]) is a probabilistic network specification language extending Kleene Algebras with Tests ([18]) with network primitives and a binary probabilistic choice operator $\oplus_\lambda, \lambda \in [0, 1]$. For the purpose of the example shown here we will not need to introduce the full syntax and semantics of ProbNetKAT, rather we will focus on a single ProbNetKAT program which we will call `cantor` and is given by:

$$\texttt{cantor} := \texttt{p}; (\texttt{dup}; \texttt{p})^* \quad \text{where} \quad \texttt{p} := \pi_0! \oplus_{1/2} \pi_1! \tag{13}$$

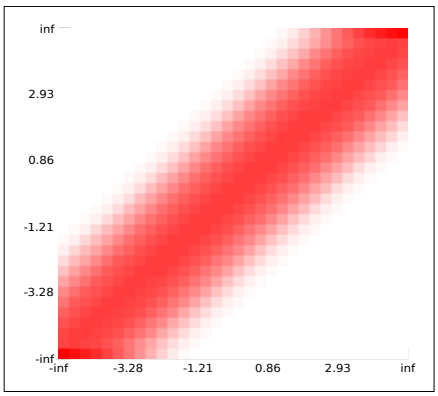

Figure 2. Log-likelihood of $f^{5,3}$

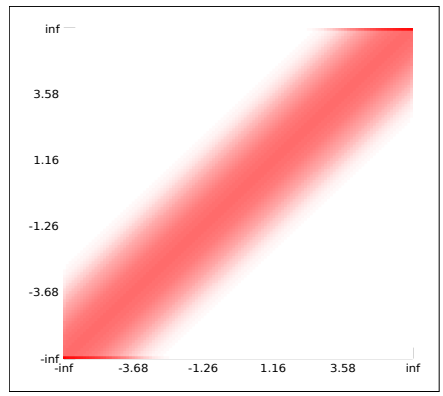

Figure 3. Log-likelihood of $f^{6,10}$

The program acts on sets of finite sequences of 0 and 1, which can be thought of as packet histories. We will write H for the set $\{0,1\}^*$ of all packet histories and H_n for the set of histories of length as most n. A ProbNetKAT program is always interpreted as a kernel $2^H \to \mathrm{G}2^H$. Programs with both dup and * revealed to be quite complex from the earliest development of the language. As we will describe, `cantor` denotes a continuous distribution and hence having a way to approximate it is crucial for practical uses of the language. The denotation of $\pi_0!$ on a single sequence $\{(\mathsf{a}_0, \ldots, \mathsf{a}_n)\}$ is:

$$[\![\pi_0!]\!](\{(\mathsf{a}_0, \ldots, \mathsf{a}_n)\}) = \delta_{\{(0, \mathsf{a}_1, \ldots, \mathsf{a}_n)\}}$$

in other words $\pi_0!$ overwrites the first entry in the sequence with `0`. Similarly, $\pi_1!$ overwrites the first entry with 1. This semantics is extended to sets of sequences in the obvious way by taking direct images. The semantics of p is thus:

$$[\![\mathsf{p}]\!](a) = 0.5\delta_{[\![\pi_0!]\!](a)} + 0.5\delta_{[\![\pi_1!]\!](a)}$$

The denotation of dup is given on singleton histories by

$$[\![\mathsf{dup}]\!](\{(\mathsf{a}_0, \ldots, \mathsf{a}_n)\}) = \delta_{\{(\mathsf{a}_0, \mathsf{a}_0, \ldots, \mathsf{a}_n)\}}$$

i.e. dup shifts the history to the right and duplicates the first entry. Again, this is extended to sets of histories by taking direct images. The sequential composition operator ; is interpreted by Kleisli composition.

The interpretation of the Kleene star is more involved, and we here describe it categorically. To avoid any confusion we will not use Kleisli arrows in this construction, i.e. all kernels will be explicitly typed as kernels. Note first that the infinite product $(2^H)^\omega$ can be defined as the limit of the ccd given by the maps $q_{n+1,n} : (2^H)^{n+1} \to (2^H)^n$ dropping the last component. By Bochner's theorem ([7]) this also holds of $\mathrm{G}((2^H)^\omega)$. Next, consider any program r. We turn 2^H into a cone for the diagram with limit $\mathrm{G}((2^H)^\omega)$ via the inductively defined maps:

$$a_1 = \eta \otimes [\![\mathsf{r}]\!] \bullet \Delta_1 : 2^H \to \mathrm{G}(2^H \times 2^H) \tag{14}$$

$$a_n = a_{n-1} \otimes [\![\mathsf{r}]\!] \bullet \Delta_n : (2^H)^n \to \mathrm{G}((2^H)^n \times 2^H) \tag{15}$$

where $\Delta_n : (2^H)^n \to (2^H)^n \times 2^H$ is the map copying the last entry. It is easy to check $q_{n+1,n} \bullet a_n \bullet a_{n-1} \bullet \ldots \bullet a_1 = a_{n-1} \bullet \ldots \bullet a_1$, and the diagram described by the morphisms $b_n := a_n \bullet \ldots \bullet a_1 : 2^H \to \mathrm{G}(2^H)^n$ makes 2^H a cone for $\varprojlim \mathrm{G}(2^H)^n$. There must therefore exist a unique morphism

$$[\![\mathsf{r}]\!]_\infty : 2^H \to \mathrm{G}\left((2^H)^\infty\right).$$

For each input, this kernel builds a distribution on the sample paths of the discrete-time stochastic processes associated with r and this input. We now define

$$[\![\mathsf{r}^*]\!] := \mathrm{G}\left(\bigcup\right) \circ [\![\mathsf{r}]\!]_\infty$$

where $\bigcup : (2^H)^\infty \to 2^H$ is the map taking infinitary unions. Since the definition above makes sense for any kernel f on 2^H, we will overload the Kleene star and put $f^* := \mathrm{G}\bigcup \circ f_\infty$. Given the input $(\mathtt{0})$, a sample path of `cantor` will draw uniformly a history of size 1, then a history of size 2 whose suffix matches the size 1 history drawn at the previous step, and so on for every integer. The distribution $[\![\mathsf{cantor}]\!](\mathtt{0})$ associates to a measurable collection of sets of histories A the probability that the union of a sample path from $(\mathtt{0})$ belongs to A. For example $[\![\mathsf{cantor}]\!](\mathtt{0})\{A \mid (\mathtt{01}) \in A\} = 1/4$, since there's a $1/4$ chance that a sample path will have drawn $(\mathtt{01})$ amongst the histories of size 2.

We start by turning 2^H into a **Krn**-object. Consider the countable directed diagram given by all injections $i_{mn} : H_m \to H_n, n > m$, then $H = \varinjlim H_n$, and it follows that $2^H = \varprojlim 2^{H_n}$ since 2^- turns colimits into limits. We know from Bochner's theorem that $\mathrm{G}2^H = \varprojlim \mathrm{G}2^{H_n}$, and we use this fact to place a canonical measure on 2^H as follows: since each 2^{H_n} is finite with cardinality $c_n := 2^{\sum_i^n 2^i} = 2^{2^{n+1}-1}$, and can thus be equipped with the uniform measure $1/c_n$, we can find a limit measure μ on 2^H with the pleasing property that for all history truncating maps $p_n : 2^H \to 2^{H_n}$, the pushforward $\mu_n := (p_n)_*\mu$ is the uniform measure on H_n. It is clear that these maps define a discretization scheme on 2^H which satisfies the condition of Theorem 35. We will now show that if $f^n \longrightarrow_s f$, then $(f^n)^* \longrightarrow_s f^*$. To prove this we need the following lemma which is interesting in its own right.

Lemma 36. *The monoidal structure of* **Krn** *is continuous for the SOT, i.e.* $f^n \longrightarrow_s f$ *and* $g^n \longrightarrow_s g$ *implies* $f^n \otimes g^n \longrightarrow_s f \otimes g$.

Theorem 37. *Under the set-up described above, for any kernel* $f : (2^H, \mu) \rightarrowtriangle (2^H, \nu)$ *we have* $(f^n)^* \longrightarrow_s f^*$

The advantage of working over finite spaces is that $(f^n)^*$ can, in principle at least, be computed for kernels defined in ProbNetKAT. Let us examine this in the case of `cantor` and of the discretization scheme $p_n : 2^H \to 2^{H_n}$.

In the case $n = 3$ the underlying Markov chain has 2^7 states, but has an interesting property which means we need not consider them all: when we compute $[\![\mathsf{p}]\!]; [\![(\mathsf{dup}; \mathsf{p})]\!]^3_\infty$, the process necessarily lands in an ergodic component of the chain consisting of the singletons of histories of length exactly 3. The reason is that once the process reaches histories of length 3 it starts randomly re-writing the histories, and with probability 1 any two histories will eventually get re-written to the same thing. Once a set of histories has decreased in cardinality by one, it can never go back, thus eventually *any* set of histories gets re-written to a single length 3 history, and then loops among length 3 singletons indefinitely. The situation is represented from the initial state (`0`) in Figure 4 where, for clarity's sake, the ergodic component is symbolized by common double-sided arrows to a new state.

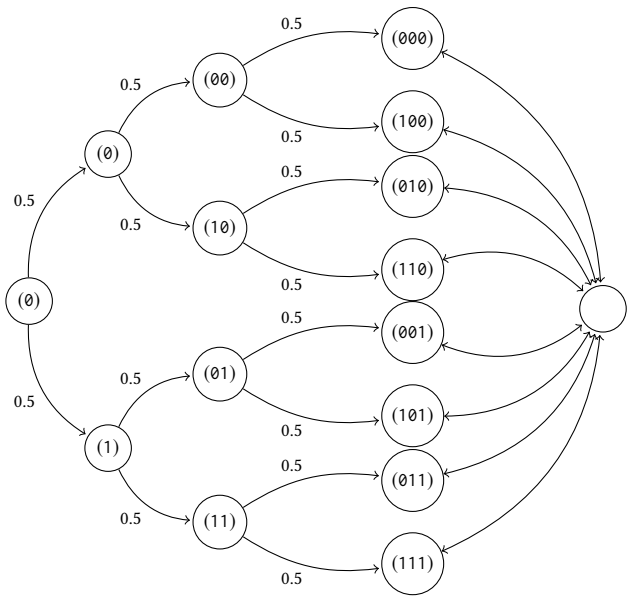

Figure 4. $[\![\mathsf{p}]\!]; [\![(\mathsf{dup}; \mathsf{p})]\!]^3_\infty$) starting at (`0`)

By post-composing with $\mathrm{G} \bigcup$ we have

$$[\![\mathsf{p}]\!]; ([\![(\mathsf{dup}; \mathsf{p})]\!]^3)^*(a)(A) = 0$$

if A does not contain all histories of length 3. We have meaningful answers to questions about histories up to length 2:

$$[\![\mathsf{p}]\!]; ([\![(\mathsf{dup}; \mathsf{p})]\!]^3)^*((\mathtt{0}))(\{A \mid (\mathtt{1}) \in A\}) = 0.5$$

$$[\![\mathsf{p}]\!]; ([\![(\mathsf{dup}; \mathsf{p})]\!]^3)^*((\mathtt{0}))(\{A \mid (\mathtt{10}) \in A\}) = 0.25$$

In other words, at $n = 3$ we have the first two steps in the construction of the Cantor distribution towards which `cantor` converges.

8 Conclusion

We have presented a framework for the exact *and* approximate semantics of first-order probabilistic programming. The semantics can be read off either in terms of kernels between measured spaces, or in terms of operators between L_p spaces. Either forms come with related involutive structures: Bayesian inversion for (measured) kernels between Standard Borel spaces, and Köthe duality for positive linear and σ-continuous operators between Banach lattices. Functorial relations between both forms can themselves be related by way of natural isomorphisms. Our main result is the convergence of general systems of finite approximants in terms of the strong operator topology (the SOT theorem). Thus, in principle, one can compute arbitrarily good approximations of the semantics of a probabilistic program of interest for any given (measurable) query. Future work may allow one to derive stronger notions of convergences given additional Lipschitz control on kernels, or to develop approximation schemes that are adapted to the measured kernel of interest. More ambitiously perhaps, one could investigate whether MCMC sampling schemes commonly used to perform approximate Bayesian inference in the context of probabilistic programming could be seen as randomized approximations of the type considered in this paper.

References

[1] N. L. Ackerman, C. E. Freer, and D. M. Roy. 2011. Noncomputable Conditional Distributions. In *LICS 2011*. 107–116.
[2] C. Aliprantis and K. Border. 1999. *Infinite dimensional analysis*. Vol. 32006. Springer.
[3] V. I. Bogachev. 2006. *Measure Theory I*. Springer.
[4] P. Chaput, V. Danos, P. Panangaden, and G. Plotkin. 2014. Approximating Markov Processes by averaging. *J. ACM* 61, 1 (Jan. 2014).
[5] K. Choo-Whan. 1972. Approximation theorems for Markov operators. *Probability Theory and Related Fields* 21, 3 (1972), 207–214.
[6] F. Clerc, V. Danos, F. Dahlqvist, and I. Garnier. 2017. Pointless learning. In *FoSSaCS*. Springer, 355–369.
[7] F. Dahlqvist, V. Danos, and I. Garnier. 2016. Giry and the Machine. *Electr. Notes Theor. Comput. Sci.* 325 (2016), 85–110.
[8] F. Dahlqvist, V. Danos, I. Garnier, and O. Kammar. 2016. Bayesian Inversion by Omega-Complete Cone Duality (Invited Paper). In *CONCUR 2016 (LIPIcs)*, Vol. 59. Schloss Dagstuhl, 1:1–1:15.
[9] V. Danos, J. Desharnais, and P. Panangaden. 2003. Conditional expectation and the approximation of labelled Markov processes. In *International Conference on Concurrency Theory*. Springer, 477–491.
[10] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. 2004. Metrics for labelled Markov processes. *TCS* 318, 3 (2004), 323–354.
[11] J. Desharnais, R. Jagadeesan, V. Gupta, and P. Panangaden. 2000. Approximating labeled Markov processes. In *LiCS 2000*. IEEE, 95–106.
[12] J. Dieudonné. 1951. Sur les espaces de Köthe. *Jour. dâĂŹAnalyse Math.* 1, 1 (1951), 81–115.
[13] N. Dunford, J. T. Schwartz, W. G Bade, and R. G. Bartle. 1971. *Linear operators I*. Wiley-interscience New York.
[14] N. Foster, D. Kozen, K. Mamouras, M. Reitblatt, and A. Silva. 2016. Probabilistic netkat. In *ESOP 2016*. Springer, 282–309.

[15] M. Giry. 1981. A Categorical Approach to Probability Theory. In *Categorical Aspects of Topology and Analysis (Lecture Notes In Math.)*. Springer-Verlag, 68–85.

[16] A. S. Kechris. 1995. *Classical descriptive set theory*. Graduate Text in Mathematics, Vol. 156. Springer.

[17] D. Kozen. 1981. Semantics of probabilistic programs. *Journal of computer and system sciences* 22, 3 (1981), 328–350.

[18] Dexter Kozen. 1997. Kleene algebra with tests. *ACM Transactions on Programming Languages and Systems (TOPLAS)* 19, 3 (1997), 427–443.

[19] D. Kozen. 2016. Kolmogorov Extension, Martingale Convergence, and Compositionality of Processes. In *LiCS 2016*. 692–699.

[20] P. Selinger. 2004. Towards a semantics for higher-order quantum computation. In *Proceedings of the 2nd International Workshop on Quantum Programming Languages, TUCS General Publication*, Vol. 33. 127–143.

[21] P. Selinger. 2007. Dagger compact closed categories and completely positive maps. *Electronic Notes in Theoretical computer science* 170 (2007), 139–163.

[22] S. Smolka, P. Kumar, N. Foster, D. Kozen, and A. Silva. 2017. Cantor meets Scott: Domain-theoretic foundations for probabilistic network programming. In *POPL 2017*.

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

[24] S. Staton, H. Yang, F. Wood, C. Heunen, and O. Kammar. 2016. Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints. In *LICS, 2016*. 525–534.

[25] D. Williams. 1991. *Probability with martingales*. Cambridge University Press.

[26] F. Wood, Jan W. van de Meent, and V. Mansinghka. 2014. A New Approach to Probabilistic Programming Inference. In *International conference on Artificial Intelligence and Statistics*. 1024–1032.

[27] A. C. Zaanen. 2012. *Introduction to operator theory in Riesz spaces*. Springer.

A Appendix

Proof of Theorem 1.
This is a consequence of the Isomorphism Theorem (Theorem 15.6 of [16]): two SB spaces are isomorphic iff they have the same cardinality. Uncountable SB spaces are thus all isomorphic to the Cantor space $2^{\mathbb{N}}$ which is the limit of the countable co-directed diagram $(2^n)_{n\in\mathbb{N}}$ with the connecting morphisms $p_{n+1,n} : 2^{n+1} \to 2^n$ truncating binary words of length $n+1$ at length n. Similarly all SB-spaces of cardinality $\aleph_0$ are isomorphic to the one-point compactification of $\mathbb{N}$, which is the limit of the countable co-directed diagram $(n)_{n\in\mathbb{N}}$ with the connecting morphisms $p_{n+1,n} : n+1 \to n, i \mapsto \min(i,n)$. The case of finite SB spaces is trivial.

■

Proof of Lemma 3.
By Dinkyn's π-λ theorem, two finite measures are equal if and only if they agree on a π-system generating the σ-algebra. Any standard Borel space admits such a countable π-system (any countable basis for a Polish topology generating the σ-algebra). Let $\{B_n\}_{n\in\mathbb{N}}$ be such a π-system. Then, for all $x \in X$, $g(x) \neq g'(x) \Leftrightarrow \exists n. g(x)(B_n) \neq g'(x)(B_n)$. Hence,

$$\begin{aligned} N(g,g') &= \cup_n\{x \in X \mid g(x)(B_n) \neq g'(x)(B_n)\} \\ &= \cup_n\{x \in X \mid ev_{B_n}(g(x)) \neq ev_{B_n}(g'(x))\} \\ &= \cup_n(ev_{B_n} \circ g - ev_{B_n} \circ g')^{-1}(\mathbb{R} \setminus \{0\}) \end{aligned}$$

By definition of the measurable structure of $\mathrm{G}(Y)$, $ev_{B_n} \circ g - ev_{B_n} \circ g'$ is measurable, hence $N(g,g')$ is also measurable.

■

Proof of Proposition 4.
We first show that if $g \sim g'$, then $h \bullet g \sim h \bullet g'$. Clearly, for any space V and any deterministic function $u : Y \to V$, $N(u\circ g, u\circ g') \subseteq N(g,g')$. By definition of the Kleisli category, $h \bullet g = m_Z \circ \mathrm{G}(h) \circ g$ and similarly for $h \bullet g'$. Taking $u = m_Z \circ \mathrm{G}(h)$, we obtain that $\mu(N(h \bullet g, h \bullet g')) \leq \mu(N(g,g'))$.

It is now enough to show that $\lambda(N(g \bullet f, g' \bullet f)) = 0$. Let us reason contrapositively. We have:

$$\begin{array}{lll} & \lambda(N(g \bullet f, g' \bullet f)) & > 0 \\ \Leftrightarrow & \int_{w\in W} 1_{N(g\bullet f, g'\bullet f)}(w)\, d\lambda & > 0 \\ \to & \int_{w\in W} \sum_{n\in\mathbb{N}} 1_{(g\bullet f)(w)(B_n)\neq(g'\bullet f)(w)(B_n)}\, d\lambda & > 0 \\ \overset{\exists n}{\to} & \int_{w\in W} 1_{(g\bullet f)(w)(B_n)\neq(g'\bullet f)(w)(B_n)}\, d\lambda & > 0 \\ \to & \int_{w\in W} |(g \bullet f)(w)(B_n) - (g' \bullet f)(w)(B_n)|\, d\lambda & > 0 \\ \Leftrightarrow & \int_{w\in W}\int_{x\in X} |g(x)(B_n) - g'(x)(B_n)|\, df(w)\, d\lambda & > 0 \\ \Leftrightarrow & \int_{x\in X} |g(x)(B_n) - g'(x)(B_n)|\, d\mu & > 0 \\ \overset{\exists X^+\subseteq X}{\to} & \int_{x\in X^+} g(x)(B_n) - g'(x)(B_n)\, d\mu & > 0 \\ \to & \int_{x\in X^+} 1_{g(-)(B_n)>g'(-)(B_n)}(x)\, d\mu & > 0 \\ \to & \int_{x\in X^+} 1_{N(g,g')}(x)\, d\mu & > 0 \end{array}$$

The last line implies $\mu(N(g,g')) > 0$, a contradiction.

■

Proof of Theorem 6.
If ϕ is ν-integrable, there exists a monotone sequence $\{\phi_n\}$ of simple functions such that $\phi_n \uparrow \phi$ and $\int_Y \phi_n d\nu \to \int_Y \phi d\nu < \infty$. By definition each $\phi_n = \sum_{i=0}^k \alpha_i 1_{B_i}$, and by unravelling the definition we have

$$\begin{aligned} \int_Y 1_{B_i} d\nu &= \nu(B_i) \\ &= \int_X f(x)(B_i) d\mu \\ &= \int_X \int_Y 1_{B_i} df(x) d\mu \\ &= \int_X (1_{B_i} \bullet f) d\mu \end{aligned}$$

From which it follows that

$$\int_Y \phi_n d\nu = \int_X \int_Y \sum_{i=0}^k \alpha_i 1_{B_i} df(x) d\mu = \int_X (\phi_n \bullet f) d\mu$$

and the result follows from the Monotone Convergence Theorem (MCT).

■

Proof of Theorem 9.
It follows by definition of $f^\dagger$ and from the disintegration theorem that

$$\int_{y\in Y} f^\dagger(y)(A) \cdot 1_B(y)\, d\nu = \gamma_f(A \times B), \tag{16}$$

from which Eq. 4 follows easily. It remains to prove that this uniquely characterizes $f^\dagger$. Let us reason contrapositively. Assume there exists $g : (Y, \nu) \to (X, \mu)$ verifying for all A, B measurable $\int_{y\in Y} g(y)(A) \cdot 1_B(y) = \gamma_f(A \times B)$ as in Eq. 16 and such that $\nu(N(f^\dagger, g)) > 0$ (assuming we take some representative of $f^\dagger$). Let $\{A_n\}_{n\in\mathbb{N}}$ be a countable π-system generating the σ-algebra of X. It is enough to test equality of measures on X on this π-system. Therefore, $N(f^\dagger, g) = \cup_n\{y \mid f^\dagger(y)(A_n) \neq g(y)(A_n)\}$. Since $\nu(N(f^\dagger, g)) > 0$, there must exist a $k \in \mathbb{N}$ such that $\nu(\{y \mid f^\dagger(y)(A_k) \neq g(y)(A_k)\}) > 0$. Therefore, $N_k^+ = \{y \mid f^\dagger(y)(A_k) > g(y)(A_k)\}$ must also have positive measure for ν. But then, $\int_{y\in Y} g(y)(A_k) \cdot 1_{N_k^+}(y) \neq \gamma_f(A_k \times N_k^+)$, a contradiction.

■

Proof of Theorem 10.
Let us first show that $(-)^\dagger$ is a functor $\mathbf{Krn} \to \mathbf{Krn}^{\mathrm{op}}$, i.e. that $\mathrm{id}^\dagger_{(X,\mu)} = \mathrm{id}_{(X,\mu)}$ and that for any $f : (X, \mu) \to (Y, \nu)$ and $g : (Y, \nu) \to (Z, \rho)$ we have $(g \bullet f)^\dagger = f^\dagger \circ g^\dagger$.

Let (X, μ) be an object of $\mathbf{Krn}$ and $\mathrm{id}_{X,\mu}$ the corresponding identity. By Th. 9, it is enough to prove, for all A, A' measurable subsets of X, that

$$\int_{x\in X} 1_A(x) \cdot id_{X,\mu}(x)(A')\, d\mu = \int_{x\in X} id_{X,\mu}(x)(A) \cdot 1_{A'}(x)\, d\mu.$$

We have:

$$\int_{x\in X} 1_A(x) \cdot id_{X,\mu}(x)(A')\, d\mu = \int_{x\in X} 1_A(x) \cdot 1_{A'}\, d\mu = \mu(A \cap A')$$

The same calculation on the right hand side of the first equation yields trivially the same result. Hence the equality is verified.

Now, on to compatibility w.r.t. composition. In sight of Th. 9, it is enough to show that for all $A \subseteq X$, $C \subseteq Z$,

$$\int_{x\in X} (g \bullet f)(x)(C) \cdot 1_A(x)\, d\mu = \int_{z\in Z} 1_C(z) \cdot (f^\dagger \bullet g^\dagger)(z)(A)\, d\rho$$

In the following, for X a measurable space, we denote by $SF(X)$ the set of simple functions over X (finite linear combinations of indicator functions of measurable sets). We will use repeatedly the monotone convergence theorem (MCT). The left hand side of the above equation can be re-written as:

$$\begin{aligned}
& \int_{x\in X} \left(\int_{y\in Y} g(y)(C)\, df(x)\right) \cdot 1_A(x)\, d\mu \\
\overset{(1)}{=}\ & \int_{x\in X} \left(\int_{y\in Y} \lim_{n\to\infty} g_n(y)\, df(x)\right) \cdot 1_A(x)\, d\mu \\
\overset{(2)}{=}\ & \int_{x\in X} \lim_{n\to\infty} \left(\int_{y\in Y} g_n(y)\, df(x)\right) \cdot 1_A(x)\, d\mu
\end{aligned}$$

where (1) is because $g_n \uparrow g(-)(C), g_n \in SF(Y)$ and (2) by monotone convergence. Note that the n-indexed family $x \mapsto \int_{y\in Y} g_n(y)\, df(x)$ is pointwise increasing. Therefore,

$$\begin{aligned}
\overset{(*)}{=}\ & \lim_{n\to\infty} \int_{x\in X} \left(\int_{y\in Y} g_n(y)\, df(x)\right) \cdot 1_A(x)\, d\mu \\
\overset{(1)}{=}\ & \lim_{n\to\infty} \int_{x\in X} \left(\textstyle\sum_{i=1}^{k_n} \alpha_i^n f(x)(C_i^n)\right) \cdot 1_A(x)\, d\mu \\
=\ & \lim_{n\to\infty} \textstyle\sum_{i=1}^{k_n} \alpha_i^n \int_{x\in X} f(x)(C_i^n) \cdot 1_A(x)\, d\mu \\
\overset{(2)}{=}\ & \lim_{n\to\infty} \textstyle\sum_{i=1}^{k_n} \alpha_i^n \int_{y\in Y} 1_{C_i^n}(y) \cdot f^\dagger(y)(A)\, d\nu \\
\overset{(*)}{=}\ & \int_{y\in Y} g(y)(C) \cdot f^\dagger(y)(A)\, d\nu \\
\overset{(3)}{=}\ & \int_{y\in Y} g(y)(C) \cdot \lim_n f_n(y)\, d\nu \\
\overset{(*)}{=}\ & \lim_n \textstyle\sum_{i=1}^{k_n} \beta_i^n \int_{y\in Y} g(y)(C) \cdot 1_{D_i^n}(y)\, d\nu \\
\overset{(2)}{=}\ & \lim_n \textstyle\sum_{i=1}^{k_n} \beta_i^n \int_{z\in Z} 1_C(z) \cdot g^\dagger(z)(D_i^n)\, d\rho \\
=\ & \lim_n \int_{z\in Z} 1_C(z) \cdot \int_{y\in Y} \textstyle\sum_{i=1}^{k_n} \beta_i^n 1_{D_i^n}(y)\, dg^\dagger(z)\, d\rho \\
\overset{(*)}{=}\ & \int_{z\in Z} 1_C(z) \cdot \int_{y\in Y} f^\dagger(y)(C)\, dg^\dagger(z)\, d\rho \\
=\ & \int_{z\in Z} 1_C(z) \cdot (f^\dagger \bullet g^\dagger)(z)(A)\, d\rho
\end{aligned}$$

where $(*)$ is by monotone convergence, (1) is because $g_n \in SF(Y)$, (2) is by Th. 9 and (3) is because $f_n \uparrow f^\dagger(-)(A), f_n \in SF(Y)$. We have proved the sought identity.

Finally let us show that $(-)^\dagger$ is involutive, i.e. that for any $f : (X, \mu) \to (Y, \nu)$, $(f^\dagger)^\dagger = f$. This follows easily by two applications of Th. 9): we have

$$\begin{aligned}
\int_{x\in X} 1_A(x) \cdot (f^\dagger)^\dagger(x)(B)\, d\mu &= \int_{y\in Y} f^\dagger(y)(A) \cdot 1_B(y)\, d\nu \\
&= \int_{x\in X} 1_A(x) \cdot f(x)(B)\, d\mu;
\end{aligned}$$

and since adjoints are unique, $f = (f^\dagger)^\dagger$.

The fact that $(f \otimes g)^\dagger = f^\dagger \otimes g^\dagger$ follows immediately from the definitions and the property of disintegrations given by Th. 9. The fact that the associator, unitors and braiding transformations are unitary follows immediately from the fact that they are deterministic isomorphisms and Th. 7.

■

Proof of Proposition 20.
Let $B \subseteq Y$ be a measurable set. By definition, we have $(f \bullet \rho)(B) = \int_X ev_B \circ f\, d\rho$ where we recall that $ev_B : \mathrm{G}(X) \to \mathbb{R}_+$ is the evaluation morphism. Let $\{f_n^B\}_{n\in\mathbb{N}}$ be an increasing chain of simple functions converging pointwise to $ev_B \circ f$ such that for each n, $f_n^B = \sum_{i=1}^{k_n} \alpha_i^n 1_{A_i^n}$ with $\alpha_i^n \geq 0$. By the MCT,

$$(f \bullet \rho)(B) = \lim_n \int_X f_n^B\, d\rho = \lim_n \sum_{i=1}^{k_n} \alpha_i^n \rho(A_i^n).$$

Similarly,

$$\nu(B) = (f \bullet \mu)(B) = \lim_n \int_X f_n^B\, d\mu = \lim_n \sum_{i=1}^{k_n} \alpha_i^n \mu(A_i^n).$$

Notice that since the integral is linear and the sequence $\{f_n^B\}_n$ is increasing, the sequences $\{\int_X f_n^B\, d\rho\}_n$ and $\{\int_X f_n^B\, d\mu\}_n$ are also increasing. Assume $\nu(B) = 0$. Then for all n, $\int_X f_n^B\, d\mu = 0$. We deduce that for all n, for all $1 \leq i \leq k_n$, either $\alpha_i^n = 0$ or $\mu(A_i^n) = 0$. Using that $\rho \ll \mu$, we deduce that for all

$1 \leq i \leq k_n$, either $\alpha_i^n = 0$ or $\rho(A_i^n) = 0$, from which we conclude that for all n, $\int_X f_n^B \, d\rho = 0$ and finally, $(f \bullet \rho)(B) = 0$. Hence, $f \bullet \rho \ll \nu$.

■

Proof of Theorem 21.
We start by proving the following Lemma

Lemma 38. *For any $f : (X, \mu) \rightarrowtriangle (Y, \nu)$, $\phi \in \mathrm{L}_1(Y, \nu)$, and $B_X \subseteq X$ measurable*

$$\int_{B_X} \left(\int_Y \phi df(x) \right) d\mu = \int_Y \phi(y) f^\dagger(y)(B_X) \, d\nu$$

Proof. We start by showing the equation on characteristic functions. If B_Y is measurable in Y, we have

$$\begin{aligned} \int_{B_X} \left(\int_Y 1_{B_Y} df(x) \right) d\mu &= \int_{B_X} f(x)(B_Y) \, d\mu \\ &= \int_Y 1_{B_Y}(y) f^\dagger(y)(B_X) \, d\nu \quad \text{Eq. (4)} \end{aligned}$$

Since ϕ is measurable and integrable, there exists a sequence $\phi_n \uparrow \phi$ of simple functions such that $\lim_n \int_Y \phi_n \, d\nu < \infty$, and the results follows by the linearity of integration and the MCT. □

We can now prove the naturality of rn. Let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ be a morphism in **Krn**; we have on the one hand

$$\begin{aligned} \mathsf{rn}_{(Y,\mu)} \circ \mathcal{M}^{\ll \cdot}(f)(\rho)(y) &= \mathsf{rn}_{(Y,\nu)}\left(\int_X f(x)(-) \, d\rho\right)(y) \\ &= \frac{d \int_X f(x)(-) d\rho}{d\nu}(y) \qquad (*) \end{aligned}$$

and on the other

$$\begin{aligned} \mathsf{T}_1(f^\dagger) \circ \mathsf{rn}_{(X,\mu)}(\rho)(y) &= \mathsf{T}_p(f^\dagger)\left(\frac{d\rho}{d\mu}\right)(y) \\ &= \int_X \frac{d\rho}{d\mu} \, df^\dagger(y) \qquad (**) \end{aligned}$$

To show the equality of these two maps in $\mathrm{L}_1(Y, \nu)$ it is enough to show that they are equal ν-a.e. To see this, we show that $(**)$ satisfies the condition to be the Radon-Nikodym derivative $(*)$. Let B_Y be a measurable subset of Y. We have from the well-known property of Radon-Nikodym derivatives:

$$\int_{B_Y} \frac{d \int_X f(x)(-) \, d\rho}{d\nu} d\nu = \int_{x \in X} f(x)(B_Y) \, d\rho$$

Moreover, we have

$$\begin{aligned} \int_{B_Y} \int_X \frac{d\rho}{d\mu} \, df^\dagger(y) \, d\nu &\overset{(1)}{=} \int_{x \in X} \frac{d\rho}{d\mu}(x) f(x)(B_Y) \, d\mu \\ &\overset{(2)}{=} \int_{x \in X} f(x)(B) \, d\rho \end{aligned}$$

where (1) is by Lemma 38 and (2) is a well-known property of Radon-Nikodym derivatives.

■

Proof of Theorem 22.
We start with the following elementary lemma.

Lemma 39. *If $\psi, \phi \in \mathrm{L}_1(X, \mu)$ then*

$$\int_X \psi\phi \, d\mu = \int_X \psi \, d(\mathsf{mr}_{(M,\mu)}\phi)$$

Proof. The proof of naturality now follows easily: it is enough to show the equality in the case where $\psi = 1_{B_X}$ for a measurable subset B_X of X, and the result then extends to all measurable functions by linearity of integrals and the MCT. We have

$$\begin{aligned} \int_X 1_{B_X} \phi \, d\mu = \int_{B_X} \phi d\mu &:= \mathsf{mr}_{(M,\mu)}(\phi)(B_X) \\ &= \int_X 1_{B_X} \, d(\mathsf{mr}_{(M,\mu)}(\phi)) \end{aligned}$$

□

To show naturality we now let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ be a **Krn**-morphism, $\phi \in \mathrm{L}_1(X, \mu)$ and B_Y measurable in Y

$$\begin{aligned} &\mathsf{mr}_{(Y,\nu)}\mathsf{T}_1(f^\dagger)(\phi)(B_X) \\ &= \mathsf{mr}_{(Y,\nu)}(\phi \bullet f^\dagger)(B_Y) \\ &= \int_{B_Y} \phi \bullet f^\dagger \, d\nu \\ &= \int_{B_Y} \int_X \phi df^\dagger(y) \, d\nu \\ &= \int_X f(x)(B_Y)\phi(x) \, d\mu && \text{Lemma 38} \\ &= \int_X f(x)(B_Y) \, d((\mathsf{mr}_{(X,\mu)}(\phi)) && \text{Lemma 39} \\ &= \mathcal{M}^{\ll \cdot} f \circ \mathsf{mr}_{(X,\mu)}(\phi)(B_Y) \end{aligned}$$

■

Proof of Theorem 23.
Again, we start with a simple but helpful Lemma.

Lemma 40. *Let $F \in (\mathrm{S}_\infty(X, \mu)^\sigma$ and $\phi \in \mathrm{S}_\infty(X, \mu)$, then*

$$F(\phi) = \int_X \phi \, d(\mathsf{rr}_{(X,\mu)}(F))$$

Proof. Starting with characteristic functions, let $\phi = 1_B$ for some measurable subset B of X. We then have

$$F(1_B) := \mathsf{rr}_{(X,\mu)}(F)(B) = \int_X 1_B \, d(\mathsf{rr}_{(X,\mu)}(F))$$

We can then extend the result to simple functions by linearity and then to all functions in $\mathrm{L}_\infty(X, \mu)$ by the MCT. □

To show naturality we now let $f : (X, \mu) \rightarrowtriangle (Y, \nu)$ be a **Krn**-morphism, $F \in (\mathrm{S}_\infty(X, \mu))^\sigma$ and B_Y measurable in Y. We

have

$$\begin{aligned}\mathcal{M}^{\ll\cdot} f \circ \mathsf{rr}_{(X,\mu)}(F)(B_Y) &= \int_X f(x)(B_Y)\, d(\mathsf{rr}_{(X,\mu)}(F)) \\ &= F(f(\cdot)(B_Y)) \qquad \text{Lemma 40} \\ &= F(\int_X 1_{B_Y} df(\cdot)) \\ &= F(1_{B_Y} \bullet f) \\ &= \mathsf{rr}_{(Y,\nu)}(F(- \bullet f))(B_Y) \\ &= \mathsf{rr}_{(Y,\nu)} \circ (\mathsf{T}_1 f)^\sigma(F)(B_Y)\end{aligned}$$

■

Proof of Theorem 24.
We start by showing that fr is well defined. The linearity of $\mathsf{fr}_{(X,\mu)}$ is easily checked on simple functions and extended by the CMT. Positivity is also immediate. For the σ-order continuity, let $\mu_m \uparrow \mu$, $\phi \in \mathrm{L}_\infty(X,\mu)$, and $\phi_n \uparrow \phi$ be a monotone approximation of ϕ by simple functions. We need to show that

$$\lim_{m\to\infty} \int_X \phi\, d\mu_m = \int_X \phi\, d\mu$$

For note first that the doubly indexed series $\int_X \phi_n d\mu_m$ is monotonically increasing in m, since the μ_m are monotonically increasing. Note also that the differences

$$d_{mn} := \int_X \phi_n\, d\mu_{m+1} - \int_X \phi_n\, d\mu_m$$

are monotonically increasing in n. Indeed we have

$$\left(\int_X \phi_{n+1}\, d\mu_{m+1} - \int_X \phi_{n+1}\, d\mu_m\right) - \left(\int_X \phi_n d\mu_{m+1} - \int_X \phi_n\, d\mu_m\right)$$
$$= \int_X (\phi_{n+1} - \phi_n)\, d\mu_{m+1} - \int_X (\phi_{n+1} - \phi_n)\, d\mu_m > 0$$

since the sequences ϕ_n and μ_m are monotonically increasing. Since d_{mn} is monotonically increasing in n we can apply the CMT to d_{mn} seen as a function of m w.r.t. the counting measure, i.e.

$$\lim_{n\to\infty} \sum_m^\infty d_{mn} = \sum_m^\infty \lim_{n\to\infty} d_{mn}$$

which is to say, by taking partial sums

$$\begin{aligned}\lim_{n\to\infty}\lim_{m\to\infty} \sum_{k=1}^m d_{kn} &= \lim_{n\to\infty}\lim_{m\to\infty} \int_X \phi_n\, d\mu_m \\ &= \lim_{m\to\infty} \sum_{k=1}^m \lim_{n\to\infty} d_{kn} \\ &= \lim_{m\to\infty}\lim_{n\to\infty} \int_X \phi_n\, d\mu_m\end{aligned}$$

which concludes the proof that fr is well-defined.

We now prove naturality. Let $f : (X,\mu) \to (Y,\nu)$ be a **Krn**-morphism, $\rho \in \mathcal{M}^{\ll\mu}(X)$ and $\phi \in \mathrm{S}_\infty(Y,\nu)$ we then have

$$\begin{aligned}\mathsf{fr}_{(Y,\nu)} \circ \mathcal{M}^{\ll\cdot} f(\rho)(\phi) &= \int_Y \phi\, d(\mathcal{M}^{\ll\cdot} f(\rho)(\phi)) \\ &= \int_Y (\phi \bullet f)\, d\rho \qquad \text{Theorem 6} \\ &= \mathrm{S}_\infty(f)\left(\int_Y (-)\, d\rho\right)(\phi) \\ &= \mathrm{S}_\infty(f) \circ \mathsf{fr}_{(X,\mu)}(\rho)(\phi)\end{aligned}$$

■

Proof of Theorem 25.
The fact that $\mathsf{rn}_{(X,\mu)}$ and $\mathsf{mr}_{(X,\mu)}$ are inverse of each other is just a restatement of the two well-known equalities for Radon-Nikodym derivatives:

$$\frac{d\int_- \phi\, d\mu}{d\mu} = \phi \quad \text{and} \quad \int_{B_X} \frac{d\rho}{d\mu}\, d\mu = \rho(B_X)$$

■

Proof of Theorem 26.
Let (X,μ) be a **Krn**-object, let $F \in (\mathrm{L}_\infty(X,\mu))^\sigma$ and let $\phi \in \mathrm{L}_\infty(X,\mu)$. We have

$$\mathsf{fr}_{(X,\mu)} \circ \mathsf{rr}_{(X,\mu)}(F)(\phi) = \int_X \phi\, d(\mathsf{rr}_{(X,\mu)}(F)) = F(\phi)$$

where the last equality follows from Lemma 40. Similarly, we have

$$\mathsf{rr}_{(X,\mu)} \circ \mathsf{fr}_{(X,\mu)}(\rho)(B_X) = \mathsf{fr}_{(X,\mu)}(\rho)(1_{B_X}) = \int_X 1_{B_X} d\rho = \rho(B_X)$$

■

Proof of Theorem 28.
The case $p = 1$ has been treated already, for the case of $1 < p < \infty$, see for example the proof of Theorem 4.4.1 of [3]. Finally for the case of $p = \infty$, see Proposition 3.3 of [4].

■

Proof of Proposition 30.
Note in (11) that we disintegrate p with respect to two different measures. For notational clarity let us define the endo-kernels

$$a := p_\mu^\dagger \bullet p$$
$$b := p_\nu^\dagger \bullet p$$

The kernel a associates to each $x \in X$ in a fibre $p^{-1}(\{i\})$ the measure $\pi_\mu^\dagger(i)$ supported by this fibre. In particular it is constant on each fibre, and similarly for b. We can now

compute:

$$
\begin{aligned}
&f^p(x)(A)\\
&:= b \bullet f \bullet a(x)(A)\\
&= \int_{y\in X} b(y)(A)\, d(f\bullet a)(x)\\
&\overset{(1)}{=} \sum_{i\in X'} \int_{y\in p^{-1}(i)} b(y)(A)\, d(f\bullet a)(x)\\
&\overset{(2)}{=} \sum_{i\in X'} p_\nu^\dagger(i)(A\cap p^{-1}(i)) f\bullet a(x)(p^{-1}(i))\\
&= \sum_{i\in X'} p_\nu^\dagger(i)(A\cap p^{-1}(i)) \int_{y\in X} f(y)(p^{-1}(i))\, da(x)\\
&\overset{(3)}{=} \sum_{i\in X'} p_\nu^\dagger(i)(A\cap p^{-1}(i)) \int_{y\in p^{-1}(p(x))} f(y)(p^{-1}(i))\, dp_\mu^\dagger(p(x))
\end{aligned}
$$

where (1) follows by decomposing X in fibres, (2) is because $b(y)(A)$ is constant on fibres, and (3) uses the fact that $a(x) := p_\mu^\dagger(p(x))$ is supported on the fibre of $p(x)$. We can considerably simplify the expression above. Note first that by definition of the disintegration

$$\nu(A\cap p^{-1}(i)) = p_\nu^\dagger(i)(A\cap p^{-1}(i))\nu(p^{-1}(i)) \tag{17}$$

Similarly, by definition of the disintegration, for $i, j \in X'$

$$
\begin{aligned}
&\int_{p^{-1}(j)} f(x)(p^{-1}(i))\, d\mu\\
&= \int_X 1_{p^{-1}(j)}(x) f(x)(p^{-1}(i))\, d\mu\\
&= \sum_{k\in X'} \mu(p^{-1}(k)) \int_{p^{-1}(k)} 1_{p^{-1}(j)}(x) f(x)(p^{-1}(i))\, dp_\mu^\dagger(k)\\
&= \mu(p^{-1}(j)) \int_{p^{-1}(j)} f(x)(p^{-1}(i))\, dp_\mu^\dagger(j)
\end{aligned} \tag{18}
$$

where the last step uses the fact that $p_\mu^\dagger(k)$ is supported by the fibre over k. By multiplying the LHS of (17), (18) we get

$$
\begin{aligned}
&\nu(A\cap p^{-1}(i)) \int_{p^{-1}(j)} f(x)(p^{-1}(i)) d\mu\\
&= \int_X f(x)(A\cap p^{-1}(i)) d\mu \int_{p^{-1}(j)} f(x)(p^{-1}(i))\, d\mu\\
&= \int_{p^{-1}(j)} f(x)(A\cap p^{-1}(i))\, d\mu \int_X f(x)(p^{-1}(i))\, d\mu\\
&= \mu(p^{-1}(j)) \int_{p^{-1}(j)} f(x)(A\cap p^{-1}(i))\, dp_\mu^\dagger(j) \int_X f(x)(p^{-1}(i))\, d\mu\\
&= \mu(p^{-1}(j))\nu(p^{-1}(A)) \int_{p^{-1}(j)} f(x)(A\cap p^{-1}(i))\, dp_\mu^\dagger(j)
\end{aligned} \tag{19}
$$

It now follows from (17), (18), and (19) that

$$
\begin{aligned}
&\mu(p^{-1}(j))\nu(p^{-1}(A)) \int_{p^{-1}(j)} f(x)(A\cap p^{-1}(i))\, dp_\mu^\dagger(j)\\
&= p_\nu^\dagger(i)(A\cap p^{-1}(i))\nu(p^{-1}(i))\mu(p^{-1}(j)) \int_{p^{-1}(j)} f(x)(p^{-1}(i))\, dp_\mu^\dagger(j)
\end{aligned}
$$

Which simplifies to

$$
\begin{aligned}
&\int_{p^{-1}(j)} f(x)(A\cap p^{-1}(i))\, dp_\mu^\dagger(j)\\
&= p_\nu^\dagger(i)(A\cap p^{-1}(i)) \int_{p^{-1}(j)} f(x)(p^{-1}(i))\, dp_\mu^\dagger(j)
\end{aligned} \tag{20}
$$

We can now use (20) to get

$$
\begin{aligned}
&b\bullet f\bullet a(x)(A)\\
&= \sum_{i\in X'} \int_{p^{-1}(\pi(x))} f(x)(A\cap p^{-1}(i))\, dp_\mu^\dagger(p(x))\\
&= \int_{p^{-1}(p(x))} f(x)(A)\, dp_\mu^\dagger(p(x)) = f\bullet a(x)(A)
\end{aligned} \tag{21}
$$

■

Proof of Lemma 34.
The map $f(-)(A) : X \to \mathbb{R}$ defines a random variable, and the discretization scheme defines a filtration $\sigma(p_n) \subseteq \sigma(p_{n+1})$ whose union is $\mathcal{B}_X$. Following Lemma 8 and Proposition 30 we have

$$f^n(x)(A) := f(x)(A)\bullet p_n^\dagger \bullet p_n = \mathbb{E}\left[f(x)(A) \mid \sigma(p_n)\right]$$

We thus have a sequence $f^n(-)(A)$ of random variables $X \to \mathbb{R}$ which is adapted to the filtration $\sigma(p_n), n \in \mathbb{N}$ by construction. We can now compute for any $m < n$

$$
\begin{aligned}
&\mathbb{E}\left[f^n(x)(A) \mid \sigma(p_m)\right]\\
&= f(x)(A)\bullet p_n^\dagger \bullet p_n \bullet p_m^\dagger \bullet p_m\\
&\overset{(1)}{=} f(x)(A)\bullet p_n^\dagger \bullet p_n \bullet (p_{nm}\bullet p_n)^\dagger \bullet p_m\\
&\overset{(2)}{=} f(x)(A)\bullet p_n^\dagger \bullet p_n \bullet p_n^\dagger \bullet p_{nm}^\dagger \bullet p_m\\
&\overset{(3)}{=} f(x)(A)\bullet p_m^\dagger \bullet p_m = f^m(x)(A)
\end{aligned}
$$

where (1) is by definition (**??**), (2) is by Thm (10) and (3) is by Theorem (7). We have thus shown that $f^n(-)(A)$ is a martingale for the filtration generated by the discretization scheme, and the result now follows from Lévy's upward convergence Theorem ([25, Th. 14.2]) since $f(x)(A) = \mathbb{E}\left[f(x)(A) \mid \sigma\left(\bigcup_n \sigma(p_n)\right)\right]$.

■

Proof of Theorem 35.
Let $(B_n)_{n\in\mathbb{N}}$ be a countable basis for the Borel σ-algebra of X, which we assume w.l.o.g. is closed under finite unions and intersections. It follows from Lemma 34 that for each B_n, $\lim_k f^k(x)(B_n) = f(x)(B_n)$ for all $x \in X \setminus N_n$ where $\mu(N_n) = 0$. It follows that for every $x \in X \setminus \bigcup_i N_i$

$$\lim_{k\to\infty} f^k(x)(B_n) = f(x)(B_n)$$

for all basic Borel sets B_n, and $\mu(\bigcup_i N_i) = 0$. Now we use the $\pi-\lambda$-lemma with $(B_n)_{n\in\mathbb{N}}$ as our π-system. We define

$$\mathcal{L} := \{C \mid f^n(x)(C) \to f(x)(C) \text{ for all } x \in X \setminus \cup_i N_i\}$$

and show that it is a λ-system. Clearly each $B_n \in \mathcal{L}$. Suppose $C \in \mathcal{L}$, it is then immediate that $C^c \in \mathcal{L}$. Now consider a sequence $C_i \in \mathcal{L}$ with $C_i \subseteq C_{i+1}$, and let $C_\infty := \cup_{i=1}^\infty C_i$. We

want to show that

$$\begin{aligned}\lim_n f^n(x)(C_\infty) &= \lim_n \lim_m f^n(x)(C_m)\\ &\overset{(*)}{=} \lim_m \lim_n f^n(x)(C_m)\\ &= \lim_m f(x)(C_m)\\ &= f(x)(C_\infty)\end{aligned}$$

where $(*)$ is the only step we need to justify. To show that the iterated limits can be switched, note first that since

$$\begin{aligned}&|f^n(x)(C_m) - f(x)(C_\infty)|\\ &= |f^n(x)(C_m) - f(x)(C_m) + f(x)(C_m) - f(x)(C_\infty)|\\ &\leq |f^n(x)(C_m) - f(x)(C_m)| + |f(x)(C_m) - f(x)(C_\infty)|\end{aligned}$$

since the two terms converge separately, for any $\epsilon > 0$ we can find $N > 0$ s. th. for all $m, n \geq N$, $|f^n(x)(C_m) - f(x)(C_\infty)| < \epsilon/2 + \epsilon/2 = \epsilon$. Thus $\lim_{(m,n)\to\infty} f^n(x)(C_m) = f(x)(C_\infty)$.

Now note also that for all $m_0 \in \mathbb{N}$ the sequence $f^n(x)(C_{m_0})$ converges to $f(x)(C_{m_0})$ (by definition of $\mathcal{L}$), and it is not hard to see that $f(x)(C_m)$ converges to $f(x)(C_\infty)$. Conversely for all $n_0 \in \mathbb{N}$, the sequence $f^{n_0}(x)(C_m)$ converges to $f^{n_0}(C_\infty)$ (by virtue of $f^{n_0}(x)$ being a measure). For $\epsilon/2 > 0$ we can find $N > 0$ such that for all $m, n > N$, $|f^n(x)(C_m) - f(x)(C_\infty)| < \epsilon/2$. We can also find $M > 0$ such that for all $m > M$, $|f^n(x)(C_m) - f^n(x)(C_\infty)| < \epsilon/2$. By taking the maximum of N and M it is clear that for all m, n above this maximum

$$\begin{aligned}&|f^n(x)(C_\infty) - f(x)(C_\infty)|\\ &\leq |f^n(x)(C_\infty) - f^n(x)(C_m)| + |f^n(x)(C_m) - f(x)(C_\infty)| < \epsilon\end{aligned}$$

We have thus shown that

$$\begin{aligned}\lim_{m\to\infty}\lim_{n\to\infty} f^n(x)(C_m) &= \lim_{(m,n)\to\infty} f^n(x)(C_m)\\ &= f(x)(C_\infty) = \lim_{n\to\infty}\lim_{m\to\infty} f^n(x)(C_m).\end{aligned}$$

Thus $\mathcal{L}$ is a λ-system, and it follows from the $\pi - \lambda$-lemma that $\sigma((B_n)_{n\in\mathbb{N}}) \subseteq \mathcal{L}$ which concludes the proof of pointwise almost everywhere convergence.

For the proof of L_1-convergence we start by showing that

$$\lim_{n\to\infty} \left\|S_p f^n(1_A) - S_p(1_A)\right\|_1 = 0 \tag{22}$$

for any Borel subset A. For this we use exactly the same reasoning as above. The only difference is that we need to check that

$$\lim_{n\to\infty} \int_X |f^n(x)(C_\infty) - f(x)(C_\infty)|\, d\mu = 0$$

For this we use the fact that we have just shown $f^n(x)(C_\infty) \to f(x)(C_\infty)$ pointwise almost everywhere, and that $|f^n(x)(C_\infty) - f(x)(C_\infty)| \leq 1$ with 1 μ-integrable. It follows by dominated convergence that

$$\begin{aligned}&\lim_{n\to\infty} \int_X |f^n(x)(C_\infty) - f(x)(C_\infty)|\, d\mu\\ &= \int_X \lim_{n\to\infty} |f^n(x)(C_\infty) - f(x)(C_\infty)|\, d\mu = 0\end{aligned}$$

which concludes the proof of (22). To extend the result to simple functions and then to arbitrary functions $\phi \in L_1(X, \nu)$ is routine. ■

Proof of Lemma 36.
Let $f^n, f : (X_1, \mu_1) \to (Y_1, \nu_1), g^n, g : (X_2, \mu_2) \to (Y_2, \nu_2)$, we need to show that for any $\phi \in L_1(Y_1 \times Y_2, \nu_1 \otimes \nu_2)$

$$\begin{aligned}\int_{X_1\times X_2} \Big| \int_{Y_1\times Y_2} \phi(y_1, y_2) d(f^n \otimes g^n)(x_1, x_2) - \\ \int_{Y_1\times Y_2} \phi(y_1, y_2) d(f \otimes g)(x_1, x_2) \Big|\, d\mu_1 \otimes d\mu_2 \to 0\end{aligned} \tag{23}$$

To show this it is enough to show that

$$\int_{Y_1\times Y_2} \phi(y_1, y_2)\, d(f^n \otimes g^n)(x_1, x_2) \to \int_{Y_1\times Y_2} \phi(y_1, y_2)\, d(f \otimes g)(x_1, x_2)$$

pointwise almost everywhere and that

$$\sup_n \left\| \int_{Y_1\times Y_2} \phi(y_1, y_2) d(f^n \otimes g^n)(x_1, x_2) \right\|_1 < \infty \tag{24}$$

Since in these circumstances pointwise convergence almost everywhere implies L_1-convergence.

To show (23) we proceed as usual: we start with simple functions and use monotone convergence. Let us first consider any measurable $A \subseteq Y_1 \times Y_2$, then

$$\begin{aligned}&\int_{(y_1,y_2)\in Y_1\times Y_2} 1_A(y_1, y_2)\, d(f^n \otimes g^n)(x_1, x_2)\\ &= \int_{y_1\in Y_1} g^n(x_2)(\{y_2 \mid (y_1, y_2) \in A\})\, df^n(x_1)\\ &\overset{(1)}{\to} \int_{y_1\in Y_1} g(x_2)(\{y_2 \mid (y_1, y_2) \in A\})\, df^n(x_1)\\ &= \int_{(y_1,y_2)\in Y_1\times Y_2} 1_A(y_1, y_2)\, d(f \otimes g)(x_1, x_2)\end{aligned}$$

where (1) is by assumption on f^n, g^n, Theorem 35 and dominated convergence. The result extends completely straightforwardly to all simple functions. Finally for an arbitrary $\phi \in L_1(Y_1 \times Y_2, \nu_1 \otimes \nu_2)$ we construct a monotone approximating sequence of simple functions $s_i \uparrow \phi$ and use the usual 3ϵ argument to conclude.

To show (24) we suppose w.l.o.g. that $\phi(x,y) \geq 0$ and compute using Fubini and Lemma 31

$$\begin{aligned}
&\left\| \int_{Y_1 \times Y_2} \phi(y_1,y_2)\, d(f^n \otimes g^n)(x_1,x_2) \right\|_1 \\
&= \int_{X_1}\int_{X_2}\int_{Y_1}\int_{Y_2} \phi(y_1,y_2)\, dg^n(x_2)\, df^n(x_1)\, d\mu_2\, d\mu_1 \\
&= \int_{X_1}\int_{Y_1}\int_{X_2}\int_{Y_2} \phi(y_1,y_2)\, dg^n(x_2)\, d\mu_2\, df^n(x_1)\, d\mu_1 \\
&\leq \int_{X_1}\int_{Y_1}\int_{X_2}\int_{Y_2} \phi(y_1,y_2)\, dg(x_2)\, d\mu_2\, df^n(x_1)\, d\mu_1 \\
&= \int_{X_2}\int_{Y_2}\int_{X_1}\int_{Y_1} \phi(y_1,y_2)\, df^n(x_1)\, d\mu_1\, dg(x_2)\, d\mu_2 \\
&\leq \int_{X_2}\int_{Y_2}\int_{X_1}\int_{Y_1} \phi(y_1,y_2)\, dg(x_2)\, df(x_1)\, d\mu_1\, d\mu_2 \\
&= \left\| \int_{Y_1 \times Y_2} \phi(y_1,y_2)\, d(f \otimes g)(x_1,x_2) \right\|_1 < \infty
\end{aligned}$$

The last step is by definition of ν_1, ν_2.

■

Proof of Theorem 37.
By definition of $(-)^*$ we need to show that $f^n_\infty \longrightarrow_s f_\infty$ and that post-composing with $\mathrm{G}\bigcup$ is continuous for the SOT. In fact, we can shown both by proving that composition is jointly continuous for the SOT and operators in the image of S_p. Note that in general composition is *not* jointly continuous for the SOT, but stochastic operators form a bounded set of operators and on these composition *is* jointly continuous in the SOT. Assume $g^n \longrightarrow_s g$ for kernels $g^n, g : (X,\mu) \rightarrowtriangle (Y,\nu)$ and $h^n \longrightarrow_s h$ for kernels $h^n, h : (Y,\nu) \rightarrowtriangle (Z,\rho)$. We then have for any $\phi \in \mathrm{L}_1(Z,\rho)$:

$$\begin{aligned}
&\|\mathrm{S}_1 g^n(\mathrm{S}_1 h^n(\phi)) - \mathrm{S}_1 g(\mathrm{S}_1 h(\phi))\| \\
&= \|(\mathrm{S}_1 g^n - \mathrm{S}_1 g)(\mathrm{S}_1 h(\phi)) + (\mathrm{S}_1 g^n(\mathrm{S}_1 h^n(\phi) - \mathrm{S}_1 h(\phi)))\| \\
&\leq \|(\mathrm{S}_1 g^n - \mathrm{S}_1 g)(\mathrm{S}_1 h(\phi))\| + \|(\mathrm{S}_1 g^n(\mathrm{S}_1 h^n(\phi) - \mathrm{S}_1 h(\phi)))\| \\
&\leq \|\mathrm{S}_1 g^n - \mathrm{S}_1 g)(\mathrm{S}_1(h(\phi))\| + \|\mathrm{S}_1 g^n\|\, \|\mathrm{S}_1 h^n(\phi) - \mathrm{S}_1 h(\phi)\|
\end{aligned}$$

where the last step follows from Hölder's inequality and the fact that $\mathrm{S}_1 g^n$, being a stochastic operator, is bounded.

Continuity under post-composition by $\mathrm{G}\bigcup$ now follows easily. For the continuity of the $(-)_\infty$ construction, we first work inductively on the construction of the maps a_k defined in (14) and (15). We denote by a^n_k the kernels generated by f^n and a_k those generated by f, $k \in \mathbb{N}$. By definition (14) $a^n_1 = \eta \otimes f^n \bullet \Delta_1$ and it follows from Lemma 36 and SOT continuity of composition that $a^n_1 \longrightarrow_s a_1$. Now assuming that $a^n_{k-1} \longrightarrow_s a_{k-1}$, the exact same argument show that $a^n_k \longrightarrow_s a_k$. Finally, the joint continuity of composition gives

$$b^n_k := a^n_k \bullet \ldots \bullet a^n_1 \longrightarrow_s b_k := a_k \bullet \ldots \bullet a_1 \tag{25}$$

for every $k \in \mathbb{N}$.

Having shown that the maps b^n_k defining f^n_∞ converge to the maps b_k defining f_∞ we now show that $\left\|\mathrm{S}_p f^n_\infty(\phi) - \mathrm{S}_p f_\infty(\phi)\right\|_1$ converges to 0 for any $\phi \in \mathrm{L}_1((2^H)^\omega, \mu^\omega)$. As usual we start with simple functions: let $\phi = \sum_i \alpha_i 1_{B_i}$ be a simple function. For any B_i we have

$$B_i \subseteq \prod_k^\infty \pi_k[B_i] \Rightarrow \mu^\infty(B_i) \leq \prod_k^\infty \mu(\pi_k)$$

and thus there can only exist finitely many indices for which the corresponding projection of B_i has not got full measure. Since the simple function is a finite sum, this means that ϕ is defined by measurable sets with non-trivial measure on only a finite set of coordinates. Let N be the largest of these coordinates, we then have:

$$\begin{aligned}
&\left\|\mathrm{S}_p f^n_\infty(\phi) - Lpfopf_\infty(\phi)\right\|_1 \\
&= \int_{2^H} \Big| \int_{(2^H)^\omega} \phi\, df^n_\infty(x) - \int_{(2^H)^\omega} \phi\, df_\infty(x) \Big| d\mu \\
&= \int_{2^H} \Big| \int_{(2^H)^N} \phi\, db^n_N(x) - \int_{(2^H)^N} \phi\, db(x) \Big|\, d\mu
\end{aligned}$$

and the result follows from (25). Extending to arbitrary maps is completely straightforward.

■