THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Bayesian Inversion by Omega-Complete Cone Duality (Invited Paper)

OPEN ACCESS

Bayesian Inversion by ω-Complete Cone Duality[*]

Fredrik Dahlqvist[1], Vincent Danos[2], Ilias Garnier[3], and Ohad Kammar[4]

1   University College London
     `f.dahlqvist@ucl.ac.uk`

2   Ecole Normale Supérieure
     `vincent.danos@ens.fr`

3   University of Edinburgh
     `igarnier@inf.ed.ac.uk`

4   University of Oxford
     `ohad.kammar@cl.cam.ac.uk`

Abstract

The process of inverting Markov kernels relates to the important subject of Bayesian modelling and learning. In fact, Bayesian update is exactly kernel inversion. In this paper, we investigate how and when Markov kernels (aka stochastic relations, or probabilistic mappings, or simply kernels) can be inverted. We address the question both directly on the category of measurable spaces, and indirectly by interpreting kernels as Markov operators:

- For the direct option, we introduce a typed version of the category of Markov kernels and use the so-called 'disintegration of measures'. Here, one has to specialise to measurable spaces borne from a simple class of topological spaces -e.g. Polish spaces (other choices are possible). Our method and result greatly simplify a recent development in Ref. [4].

- For the operator option, we use a cone version of the category of Markov operators (kernels seen as predicate transformers). That is to say, our linear operators are not just continuous, but are required to satisfy the stronger condition of being ω-chain-continuous.[1] Prior work shows that one obtains an adjunction in the form of a pair of contravariant and inverse functors between the categories of L_1- and L_∞-cones [3]. Inversion, seen through the operator prism, is just adjunction.[2] No topological assumption is needed.

- We show that both categories (Markov kernels and ω-chain-continuous Markov operators) are related by a family of contravariant functors T_p for $1 \leq p \leq \infty$. The T_p's are Kleisli extensions of (duals of) conditional expectation functors introduced in Ref. [3].

- With this bridge in place, we can prove that both notions of inversion agree when both defined: if f is a kernel, and $f^\dagger$ its direct inverse, then $\mathsf{T}_\infty(f)^\dagger = \mathsf{T}_1(f^\dagger)$.

---

[*] This work was funded partly through the RULE/320823 ERC project.
[1] This stronger continuity condition derives from domain-theoretic ideas and is due to Selinger [14].
[2] A similar L_p/L_q duality result also exists [2], and we will investigate it in later work.

1   Introduction

Before we get in the technicalities, we review informally the central notions of interest: kernels, kernels as operators, and Bayesian update. We also take the opportunity to introduce some of the notations used in the remainder of the paper.

A kernel is a (measurable) map from some measurable space X to the set of probability distributions on another measurable space Y. Probability distributions have to be equipped with a set of mesasurables for this to make sense. We will write $X \to \mathsf{G}Y$ or equivalently $X \rightarrowtail Y$ for the type of such X, Y kernels, where $\mathsf{G}X$ is the set of probability distributions over X. With the appropriate constructions, G is a monad over **Mes**, the category of measurable spaces and measurable maps [9].

Kernels are often used as models of probabilistic behaviour. If $X = Y$ is finite, then this is but the familiar notion of probabilistic state machine (or finite discrete-time Markov chain), where one jumps probabilistically from one state to the next with no dependence on the past. We say a kernel $f : X \to \mathsf{G}Y$ is deterministic[3] if it factorises in **Mes** as $f = \delta_Y \circ f'_d$ for some $f_d : X \to Y$, where $\delta_X : X \rightarrowtail X$ sends x to δ_x the Dirac measure at x. Intuively determinism means that f allows only one possible jump at each x in X. In particular, δ_X is a deterministic kernel itself (with an identical jump).

One can also use a kernel as a family of probabilities over Y parameterised by X. Each probability in the range of the kernel can be thought as a competing description of a hidden true probability. Given an additional probability p on the parameter space X, a *prior*, one can specify how much 'trust' one has in any particular description offered. This is the Bayesian model of quantification of uncertain probabilities: the prior describes our beliefs and 'Bayesian update' is a process by which new data (minted by the true random source) can be incorporated to modify the prior, and update our beliefs. The hope is that in the long run the successive priors will take us closer to the truth.
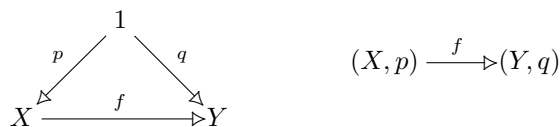
Kernels as operators

Yet another standpoint on kernels is to look at them as linear maps. Indeed, a finite $f : X \to \mathsf{G}Y$ can be seen as a transition matrix $\mathsf{T}(f)$ of type $X \times Y$ with the x, y-entry specifying the probability that being at x in X one jumps to y in Y.[4] Thus, one can think of $f : X \to \mathsf{G}Y$ as a linear map from the free vector space over Y to that over X. Clearly $\mathsf{T}(\delta_X) = I_X$. In particular, a probability p over X, ie a map from $1 \to \mathsf{G}X$, is a matrix $\mathsf{T}(p)$ of type $1 \times X$ (a row vector). This new standpoint gives a ready access to the key operation of kernel composition, written $\circ_{\mathsf{G}}$. Because G is a monad (known as the Giry monad), one knows how to compose kernels for general reasons using the so-called *Kleisli composition* (see Section 2). In the operator interpretation, Kleisli composition is just plain matrix multiplication. E.g. the composition $f \circ_{\mathsf{G}} p$ of p and f is represented as $\mathsf{T}(p)\mathsf{T}(f)$, a new row vector of type $1 \times Y$. The T (contravariant) functor can be extended to arbitrary measurable kernels, using the machinery of Banach cones of real functions (see Section 4.4).[5]

---

[3]   Or a deterministic map, following the terminology of Lawvere in his seminar handout notes on the "category of probabilistic mappings" (1962).

[4]   Of course the transpose representation is also possible as will be clear in the duality of the operator interpretation.

[5]   This predicate transformer view was first championed by Kozen for probabilistic programs [12].

Importantly, we will deal not with 'naked' Markov kernels, but with 'typed' ones of the form:[6]

$$
\begin{array}{ccc}
 & 1 & \\
{}^{p}\swarrow & & \searrow{}^{q} \\
X & \xrightarrow{\ \ f\ \ } & Y
\end{array}
\qquad\qquad
(X,p) \xrightarrow{\ \ f\ \ } (Y,q)
$$

where the triangle (above, left) is drawn in the Kleisli category of G in **Mes** and assumed to commute. The simpler diagram (above, right) is just a more compact notation for the same. (Recall that we use blocky arrows to remind us that these are Kleisli arrows.) Either diagram means that in addition to f, we are given probabilities p on X, q on Y with $f \circ_{\mathsf{G}} p = q$.

This category of typed kernels has a natural subcategory where f is deterministic. The typing simply amounts to saying that q is the push-forward of p along g.[7] E.g. $\delta_X :$ $(X,p) \to (X,p)$ for any p in GX. This subcategory is (equivalent to) the familiar category of probabilistic triples and measurable measure-preserving maps.[8]

Bayesian update or inversion

Our main question is as folllows. Given a typed kernel f, we wish to build and characterise a 'weak' inverse $f^{\dagger}$:

$$
(X,p) \underset{f^{\dagger}}{\overset{f}{\rightleftarrows}} (Y,q)
$$
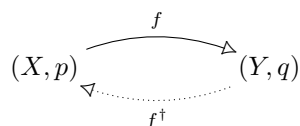
In the Bayesian world, $f^{\dagger}$ should be the update map we mentioned above. Given a data input y in Y, this map returns a *posterior* $f^{\dagger}(y)$ which represents our updated set of beliefs. It is therefore central to the theory to obtain good and general descriptions of $f^{\dagger}$. We will access such descriptions of $f^{\dagger}$ following two routes. The direct route uses the non-trivial notion of disintegration (aka regular conditional probabilities) which solves the inversion problem in the special case of a deterministic f. A clever construction based on couplings allows one to generalise it to any kernel. This is done in Section 3.3 and follows the construction of Ref. [4] while being markedly simpler. The other route goes the operator way. As alluded to in the abstract, we use a domain theoretic variant of the usual interpretation which carries a perfect duality (which one does not have in the standard interpretation). Then, as the notation suggests, we can just define $\mathsf{T}(f)^{\dagger}$ as the adjoint of $\mathsf{T}(f)$ (this is done in Section 4). Bayesian inversion is cone duality! Whence the title. We then show that through our functorial bridge, T (really a family T_p), both routes agree (Section 5).

The paper is almost self-contained. The only pieces of mathematics borrowed are the disintegration and the cone duality results and some of the most basic definitions. We now turn to the technical preliminaries.

---

[6] In category-theoretical words, our arrows are in the 'category under 1' of the Kleisli of G.
[7] Because in this case $f \circ_{\mathsf{G}} p = (\delta \circ f_d) \circ_{\mathsf{G}} p = \mu \circ \mathsf{G}(\delta \circ f_d) \circ p = \mu \circ \mathsf{G}(\delta) \circ \mathsf{G}(f_d) \circ p = \mathsf{G}(f_d) \circ p$; as follows from the monad structure equation $\mu \circ \mathsf{G}(\delta) = I$.
[8] Richer categories of kernels were considered before; for instance, to obtain a notion of almost sure bisimilarity on labelled kernels (aka labelled Markov processes) Ref. [6, Section 7] equips kernels with ideals of negligibles.

2   Preliminaries

Measurable and Polish spaces

We refer to Ref. [1] for the definitions of measurable spaces and maps, and Ref. [11] for an introduction to the theory of Polish spaces (completely metrisable and separable topological spaces). Where convenient, we will denote measurable spaces and related structures by their underlying set. If X is a set, (Y, Λ) a measurable space and $f : X \to Y$ a function, we denote by $\sigma(f)$ its *initial σ-algebra*, which is the smallest σ-algebra that makes it measurable. As seen earlier, the category of measurable spaces and measurable maps is denoted by **Mes**, and that of Polish spaces and continuous maps by **Pol**. There is a functor $\mathcal{B} : \textbf{Pol} \to \textbf{Mes}$ associating any Polish space to the measurable space with same underlying set equipped with the 'Borel' σ-algebra (generated by open sets), and interpreting continuous maps as measurable ones. Measurable spaces in the range of $\mathcal{B}$ are the *standard Borel spaces*.

A *measure p* on a measurable space (X, Σ) is a set function $\Sigma \to \mathbb{R}$ which is σ-additive and such that $p(\emptyset) = 0$. One says p is a *finite measure* whenever $p(X) < \infty$, and a probability measure if $p(X) = 1$. A property holds *p-almost surely* (p-a.s) if its negation holds on a set of measure 0. A *measure space* is a triple (X, Σ, p) such that (X, Σ) is a measurable space and p is a finite measure on (X, Σ). We denote by $p|_\Lambda$ the restriction of p to a sub-σ-algebra $\Lambda \subseteq \Sigma$. A Borel measurable real-valued function $f : (X, \Sigma, p) \to \mathbb{R}$ is called *integrable* if $\int_X |f|\, dp < \infty$.

Radon-Nikodym and conditional expectations

Let (X, Σ) be some measurable space. For p, q finite measures, we say that p is absolutely continuous with respect to q if for all B measurable, $q(B) = 0$ implies $p(B) = 0$. This will be denoted by $p \ll q$. The Radon-Nikodym theorem tells us that we can express p in terms of its *derivative* with respect to q:

▶ **Theorem 1** (Radon-Nikodym)**.** *If $p \ll q$ there exists a q-a.s. unique positive integrable function denoted by $\frac{dp}{dq} : (X, \Sigma, q) \to \mathbb{R}$ such that $p = B \mapsto \int_B \frac{dp}{dq} dq$.*

The function $\frac{dp}{dq}$ is called the Radon-Nikodym derivative of p with respect to q.

Let us denote $f \cdot p = B \in \Lambda \mapsto \int_B f\, dp$. Clearly $f \cdot p \ll p$. The following two identities follow from Theorem 1: (i) $\frac{df \cdot p}{dp} = f$ (ii) $\frac{dp}{dq} \cdot q = p$. We refer the reader to [1] for further facts about Radon-Nikodym derivatives. Conditional expectations can be implemented in terms of Radon-Nikodym derivatives.

▶ **Definition 2** (Conditional expectation ([10]))**.** Let (X, Σ, p) be a measure space and let $\Lambda \subseteq \Sigma$ be a sub-σ-algebra. The *conditional expectation* of an integrable function $f : (X, \Sigma, p) \to \mathbb{R}$ with respect to Λ is the p-a.s. unique integrable function $\mathbb{E}\left[f \mid \Lambda\right] : (X, \Lambda, p|_\Lambda) \to \mathbb{R}$ that verifies for all $B \in \Lambda$ the identity $\int_B \mathbb{E}\left[f \mid \Lambda\right] dp = \int_B f\, dp$.

Theorem 1 implies the existence of conditional expectations: letting $f \cdot p = B \in \Lambda \mapsto \int_B f\, dp$, we have for all $B \in \Lambda$ the characteristic identity $\int_B \frac{df \cdot p}{dp|_\Lambda} dp|_\Lambda = \int_B f \cdot p = \int_B f dp$.

Probability functors

The endofunctor $\mathsf{G} : \textbf{Mes} \to \textbf{Mes}$ associates to any measurable space X the set of all probability measures on X with the smallest σ-algebra that makes the evaluation functions $ev_B : \mathsf{G}(X) \to \mathbb{R} = p \mapsto p(B)$ measurable, for B a measurable set in X. If $f : X \to Y$ is measurable, the action of the functor is defined by $\mathsf{G}(f)(P) = P \circ f^{-1}$. This functor can be

endowed with the structure of the Giry monad $(\mathsf{G}, \mu, \delta)$. The multiplication $\mu : \mathsf{G}^2 \Rightarrow \mathsf{G}$ is defined at a component X by $\mu_X(P)(B) = \int_{\mathsf{G}(X)} ev_B dP$ while the unit $\delta : Id \Rightarrow \mathsf{G}$ is defined at X by $\delta_X(x) = \delta_x$, where $\delta_x(B) = 1$ if and only if $x \in B$.

The Kleisli category of G will be denoted by $\mathcal{K}\ell$. It has the same objects as **Mes**. For all X, Y measureable spaces, a Kleisli arrow $f : X \rightarrow Y$ in $\mathcal{K}\ell$ is a kernel $f : X \to \mathsf{G}(Y)$ in **Mes**. The Kleisli composition of the kernel $f : X \rightarrow Y$ with $g : Y \rightarrow Z$ is given by $g \circ_{\mathsf{G}} f = \mu_Y \circ \mathsf{G}(g) \circ f$.

3 Bayesian inversion

Let D be a space representing some space of *data* and let $t \in \mathsf{G}(D)$ be the *truth*, a unknown probability measure that we wish to discover by sampling repeatedly from it. In order to make this search analytically or computationally tractable, or more generally to reflect some additional knowledge or assumptions held about the truth, one might wish to parameterise the search through a space H of *parameters* and a measurable *likelihood function* $f : H \rightarrow D$. The uncertainty about which parameter best matches the truth is represented by a probability $p \in \mathsf{G}(H)$ called the *prior*. The composite of the two arrows $q = f \circ_{\mathsf{G}} p$ is called the *marginal likelihood*.

Bayesian inversion is the construction from these data of a *posterior map* $g : D \rightarrow H$, also called the *inference map*. Upon observing a sample $d \in D$, this inference map will produce an updated prior $g(d)$. In good cases (e.g. H and D finite, and q absolutely continuous w.r.t. t), sampling independently and identically from the truth t and iterating this Bayesian update will make the marginal likelihood converge (in some topology to be chosen carefully) to t. The key step in the above process is the construction of the posterior g, which relies crucially on *disintegrations*.

Culbertson & Sturtz give in [4] a nice categorical account of Bayesian inversion in a setting close to $\mathcal{K}\ell$. In the following, we provide a streamlined view of their work by defining a category of kernels where disintegration and Bayesian inversion admit rather elegant statements.

3.1 Categories of kernels

Let $F : \mathbf{Mes} \to \mathcal{K}\ell$ be the functor embedding **Mes** into the Kleisli category of G. It acts identically on spaces and maps measurable arrows $f : X \to Y$ to Kleisli arrows $F(f) = \delta_Y \circ f$. $1 \downarrow F$ is the category having as objects probabilities $p : 1 \rightarrow X$, denoted by (X, p), and as morphisms $f : (X, p) \rightarrow_\delta (Y, q)$ degenerate Kleisli arrows $F(f) : X \rightarrow Y$ such that $q = F(f) \circ_{\mathsf{G}} p = \mathsf{G}(f)(p)$. As said, these correspond to the usual notion of measure-preserving map. $1 \downarrow \mathcal{K}\ell$ is the category having the same objects as $1 \downarrow F$ but where arrows are non-degenerate, i.e. an arrow from (X, p) to (Y, q) as above is *any* Kleisli arrow $f : X \rightarrow Y$ such that $q = f \circ_{\mathsf{G}} p$. Clearly, $1 \downarrow F$ is a subcategory of $1 \downarrow \mathcal{K}\ell$ with the same objects (aka lluf).

The following result ensures that for an arrow $f : (X, p) \rightarrow (Y, q)$, there are p-negligibly many points jumping to q-negligible sets (it corresponds to the condition of *non-singularity* of [3]).

▶ **Lemma 3.** *If $f : (X, p) \rightarrow (Y, q)$ is an arrow in $1 \downarrow \mathcal{K}\ell$, then $f(x) \ll q$ p-a.s.*

Proof. By definition of $1 \downarrow \mathcal{K}\ell$, $q(B) = \int_X f(x)(B) \, dp$. Assume $q(B) = 0$, then having $f(x)(B) > 0$ on a set of strictly positive p-measure implies that the integral is strictly positive, yielding a contradiction.   ◀

For all objects $(X, p), (Y, q)$, let $R_{(X,p),(Y,q)}$ be the smallest equivalence relation on $\mathsf{Hom}_{1 \downarrow \mathcal{K}\ell}(X, Y)$ such that $(f, f') \in R_{(X,p),(Y,q)}$ if f and f' are p-a.s. equal.

▶ **Lemma 4.** *R defines a congruence relation on* $1 \downarrow \mathcal{K}\ell$.

Proof. We must show that for all $g : (X', p') \rightarrow (X, p)$ and all $h : (Y, q) \rightarrow (Y', q')$, $(h \circ_{\mathbf{G}} f \circ_{\mathbf{G}} g, h \circ_{\mathbf{G}} f' \circ_{\mathbf{G}} g) \in R_{(X',p'),(Y',q')}$. First, let us prove that $(f \circ_{\mathbf{G}} g, f' \circ_{\mathbf{G}} g) \in R_{(X',p'),(X,p)}$. By Lemma 3, $g(x') \ll p$ p'-a.s., hence the following equation holds for p'-almost all x':

$$(f \circ_{\mathbf{G}} g)(x') = B_Y \mapsto \int_{x \in X} f(x)(B_Y) dg(x') = B_Y \mapsto \int_{x \in X} f'(x)(B) dg(x') = (f' \circ_{\mathbf{G}} g)(x')$$

It remains to prove that $h \circ_{\mathbf{G}} f \circ_{\mathbf{G}} g$ is p'-a.s. equal to $h \circ_{\mathbf{G}} f' \circ_{\mathbf{G}} g$. We have:

$$
\begin{array}{rcll}
(h \circ_{\mathbf{G}} f \circ_{\mathbf{G}} g)(x') & = & B_{Y'} \mapsto \int_{y' \in Y'} h(y')(B_{Y'}) d(f \circ_{\mathbf{G}} g)(x') & \\
& = & B_{Y'} \mapsto \int_{y' \in Y'} h(y')(B_{Y'}) d(f' \circ_{\mathbf{G}} g)(x') & p'\text{-a.s.} \\
& = & (h \circ_{\mathbf{G}} f' \circ_{\mathbf{G}} g)(x') & p'\text{-a.s.}
\end{array}
$$

which concludes the proof.   ◀

This congruence relation allows us to consider R-equivalence classes of $1 \downarrow \mathcal{K}\ell$ arrows as proper morphisms in the corresponding quotient category (Section 2.8, [13]):

▶ **Definition 5.** The category **Krn** is the quotient category $(1 \downarrow \mathcal{K}\ell)/R$, with subcategory $\mathbf{Krn}_\delta = (1 \downarrow F)/R$.

In other terms, an arrow $f : (X, p) \twoheadrightarrow (Y, q)$ in **Krn** is an equivalence class of kernels that are p-a.s. equal.

3.2 Disintegrations

Disintegrations are also called *regular conditional probabilities* and correspond to measurable families of conditional probabilities. Working in the setting of standard Borel spaces ensures their existence, and the corresponding statement admits a particularly elegant form in **Krn**:

▶ **Theorem 6** (Disintegration, [8]). *Let X and Y be standard Borel spaces, and let $f : (X, p) \twoheadrightarrow_\delta (Y, q)$ be an arrow in $\mathbf{Krn}_\delta$. There exists a unique **Krn** arrow $f^\dagger : (Y, q) \twoheadrightarrow (X, p)$ that verifies $f^\dagger(y)(f^{-1}(\{y\})) = 1$ q-a.s.*

We call $f^\dagger$ the *disintegration* of p along f. We will show in Section 5 that disintegrations and more generally Bayesian inverses are *adjoints*, hence the use of the $-^\dagger$ notation. The last condition can be equivalently stated as the fact that $f \circ f^\dagger = id_{(Y,q)}$. In order to bridge our crisp statement of Theorem 6 with the usual measure-theoretic one, let us unfold the objects at play. If we inspect the type of the arrows $f : (X, p) \twoheadrightarrow_\delta (Y, q)$ and $f^\dagger : (Y, q) \twoheadrightarrow (X, p)$ we see that by definition of composition in $\mathcal{K}\ell$, we have the equation $q = (\mu_Y \circ \mathbf{G}(\delta_Y \circ f))(p) = \mathbf{G}(f)(p)$. The existence of the disintegration arrow $f^\dagger : (Y, q) \twoheadrightarrow (X, p)$ implies the equation $p = (\mu_Y \circ \mathbf{G}(f^\dagger))(q)$ which corresponds to

$$
\begin{array}{rcll}
p & = & B \mapsto \int_{\mathbf{G}(Y)} ev_B \, d\mathbf{G}(f^\dagger)(q) & \\
& = & B \mapsto \int_{y \in Y} f^\dagger(y)(B) \, dq & \text{(Change of variables)}
\end{array}
\tag{1}
$$

We recall that the uniqueness of $f^\dagger$ claimed in Theorem 6 is really that of a q-equivalence class of kernels. Note also that disintegrations do not in general exist in **Pol** as they need not be continuous (even when disintegrating along a continuous map).

Disintegration, as said above, is a measurable family of conditional probabilities. The sub-σ-algebras against which the conditionings are performed are encoded through the measurable map f along which the disintegration is computed. Simple calculations make explicit how conditional expectation underpins disintegration: for all $h : (X, p) \to \mathbb{R}$ integrable,

$$
\begin{aligned}
\int_X h \, dp &= \int_{y \in Y} \left(\int_X h \, df^\dagger(y) \right) dq &&\text{(Equation 1)} \\
&= \int_{y \in Y} \left(\int_{f^{-1}(y)} h \, df^\dagger(y) \right) dq &&\text{(Theorem 6)}
\end{aligned}
$$

The last equation corresponds to the usual measure-theoretic characteristic identity of disintegrations. Let us consider a measurable set $B \in \sigma(f)$ and let us apply this identity to the function $h \cdot \mathbb{1}_B$. Applying a change of variables on $q = \mathsf{G}(f)(p)$, we get:

$$
\begin{aligned}
\int_B h \, dp &= \int_{x \in X} \left(\int_{f^{-1}(f(x))} h \cdot \mathbb{1}_B \, df^\dagger(f(x)) \right) dp &&\text{(Change of variables)} \\
&= \int_{x \in B} \left(\int_{f^{-1}(f(x))} h \, df^\dagger(f(x)) \right) dp &&\text{($\mathbb{1}_B$ constant on fibers)}
\end{aligned}
$$

We recognise the characteristic identity of conditional expectations (Definition 2). This implies that the following identity holds p-almost everywhere:

$$
\mathbb{E}\left[h \mid \sigma(f)\right] = x \mapsto \int_{f^{-1}(f(x))} h \, df^\dagger(f(x)) \tag{2}
$$

3.3 Bayesian inversion

Bayesian inversion is a reformulation of the disintegration theorem where the map f is allowed to be any arrow of **Krn** (and not just a deterministic one). To formulate our *Bayesian inversion theorem* we will define two **Set**-valued functors and two natural transformations between them. The first functor is simply the functor $\mathsf{Hom}((X, p), -) : \mathbf{Krn} \to \mathbf{Set}$ for a given (X, p) in **Krn**. For notational clarity, we will abbreviate objects $(X, p), (Y, q), (Y', q')$ in **Krn** to X, Y, Y' with the understanding that they come equipped with measures p, q, q'. It is useful to explicitly write the action of $\mathsf{Hom}(X, -)$ on morphisms $g : Y \rightarrowtail Y'$. By definition $\mathsf{Hom}(X, -)(g) : \mathsf{Hom}(X, Y) \to \mathsf{Hom}(X, Y')$ maps $f \in \mathsf{Hom}(X, Y)$ to the kernel:

$$
\mathsf{Hom}(X, -)(g)(f) \triangleq g \circ_{\mathbf{G}} f : X \rightarrowtail Y' \text{ defined by } (g \circ_{\mathbf{G}} f)(x)(B_{Y'}) = \int_{y \in Y} g(y)(B_{Y'}) \, df(x)
$$

Given X in **Krn**, our second functor $\Gamma(X, -) : \mathbf{Krn} \to \mathbf{Set}$ is defined on objects as follows: we define $\Gamma(X, Y) \subseteq \mathsf{G}(X \times Y)$ to be the set of *couplings* of p and q, corresponding to measures γ such that $\mathsf{G}(\pi_X)(\gamma) = p$ and $\mathsf{G}(\pi_Y)(\gamma) = q$. Couplings corresponds to elements γ such that the following diagram commutes in $\mathcal{K}\ell$:

(3)

On morphisms $g : Y \rightarrowtail Y'$, $\Gamma(X, g)$ is defined as the map $\Gamma(X, g) = \otimes \circ_{\mathbf{G}} (\delta \times g) \circ_{\mathbf{G}} -$ such that $\Gamma(X, g)(\gamma)$ is the composite

$$
1 \xrightarrow{\;\gamma\;} X \times Y \xrightarrow{\;\otimes \circ (\delta_X \times g)\;} X \times Y'
$$

where $\otimes : \mathsf{G}X \times \mathsf{G}Y' \to \mathsf{G}(X \times Y')$ is the product measure bifunctor and δ_X is the Giry unit at X. By unravelling the definitions we get

$$\Gamma(X, g)(\gamma)(B_X \times B_{Y'}) = \int_{(x,y) \in X \times Y} \delta_X(x)(B_X) \cdot g(y)(B_{Y'}) \, d\gamma$$

The proof that $\Gamma(X, -)$ commutes with composition follows from the disintegration theorem. We now define a transformation $\alpha^X : \mathsf{Hom}(X, -) \to \Gamma(X, -)$ defined at Y by

$$\alpha^X_Y(f)(B_X \times B_Y) = \int_{x \in B_X} f(x)(B_Y) \, dp$$

▶ **Proposition 7.** α^X *is natural*

Proof. Let $g : (Y, q) \to (Y', q')$, we calculate

$$
\begin{aligned}
\Gamma(X, g)(\alpha^X_Y(f))(B_X \times B_{Y'}) &\overset{(1)}{=} \int_{(x,y) \in X \times Y} \delta_X(x)(B_X) g(y)(B_{Y'}) \, d\alpha^X_Y(f) \\
&\overset{(2)}{=} \int_{x \in X} \delta_X(x)(B_X) \left(\int_{y \in Y} g(y)(B_{Y'}) df(x) \right) \, dp \\
&= \int_{x \in B_X} \left(\int_{y \in Y} g(y)(B_{Y'}) \, df(x) \right) \, dp \\
&\overset{(3)}{=} \alpha^X_{Y'}(\mathsf{Hom}(X, g)(f))(B_X \times B_{Y'})
\end{aligned}
$$

where (1) follows from the definition of Γ, (2) follows from the fact that α^X_Y constructs a coupling in an explicitly disintegrated form, and (3) follows from the definition of α^X and $\mathsf{Hom}(X, -)$. ◀

Our second natural transformation goes in the opposite direction and is given by the disintegration along the first projection (which exists by Theorem 6), i.e. we define $D^X : \Gamma(X, -) \to \mathsf{Hom}(X, -)$ at Y in **Krn** by:

$$D^X_Y(\gamma) = \mathsf{G}(\pi_Y) \circ \pi^\dagger_X, \text{ such that } \gamma(B_X \times B_Y) = \int_{B_X} D^X_Y(\gamma)(x)(B_Y) dp(dx)$$

▶ **Proposition 8.** D^X *is natural.*

Proof. Let $g : (Y, q) \to (Y', q')$ and $\gamma \in \Gamma(X, Y)$. For notational clarity, for any $h \in \mathsf{Hom}(X, Y)$ let us write $\tilde{h} = \mathsf{Hom}(X, -)(g)(h)$, and let us define $f \triangleq D^X_Y(\gamma)$. We now calculate:

$$
\begin{aligned}
\Gamma(X, g)(\gamma)(B_X \times B_{Y'}) &\overset{(1)}{=} \int_{(x,y) \in X \times Y} \delta_X(x)(B_X) g(y)(B_{Y'}) \, d\gamma \\
&\overset{(2)}{=} \int_{x \in X} \delta_X(x)(B_X) \left(\int_{y \in Y} g(y)(B_{Y'}) \, df(x) \right) \, dp \\
&= \int_{x \in B_X} \left(\int_{y \in Y} g(y)(B_{Y'}) \, df(x) \right) \, dp \\
&\overset{(3)}{=} \int_{x \in B_X} \tilde{f}(x)(B_{Y'}) \, dp
\end{aligned}
$$

where (1) follows by definition of $\Gamma(X, -)$, (2) is by definition of f and the Disintegration Theorem 6, and (3) by definition of $\mathsf{Hom}(X, -)$. It follows immediately that $\tilde{f}$ factors through the disintegration of $\Gamma(X, g)(\gamma)$ along the first projection, i.e. that

$$D^X_{Y'}(\Gamma(X, g)(\gamma)) = \mathsf{Hom}(X, g)(D^X_Y(\gamma))$$

◀

▶ **Theorem 9** (Bayesian Inversion Theorem). *There exists a bijection:*

$$-^\dagger : \mathsf{Hom}_{\mathbf{Krn}}((X, p), (Y, q)) \to \mathsf{Hom}_{\mathbf{Krn}}((Y, q), (X, p))$$

Proof. It is immediate from the definitions above that $\alpha_Y^X : \mathsf{Hom}(X, Y) \to \Gamma(X, Y)$ and $D_Y^X : \Gamma(X, Y) \to \mathsf{Hom}(X, Y)$ are inverse of one another, and thus bijective. Moreover, it is also clear that the permutation map $\mathsf{G}(\pi_2 \times \pi_1) : \Gamma(X, Y) \to \Gamma(Y, X)$ being its own inverse is a bijection. It follows that $\mathsf{Hom}(X, Y) \simeq \Gamma(X, Y) \simeq \Gamma(Y, X) \simeq \mathsf{Hom}(Y, X)$. And we can set $-^\dagger = D_X^Y \circ \mathsf{G}(\pi_2 \times \pi_1) \circ \alpha_Y^X$.   ◀

4 ω-complete normed cones

We recall some facts pertaining to the categories of ω-complete normed cones introduced in [14]. The main object of this section is to give a functional analytic account of kernels as operators on ω-complete normed cones, in the style of [3]. We improve on the latter by presenting the transformation from kernels to operators functorially. In Section 5, we will use the machinery developed here to interpret Bayesian inversion in this domain-theoretic setting. We first recall some general definitions about ω-complete normed cones and the associated category $\omega\mathbf{CC}$. We then concentrate on the duality existing between the subcategories of cones of integrable and bounded functions. Proofs not provided here can be found in Ref. [3].

4.1   Basic definitions

Cones are axiomatisations of the positive elements of (real) vector spaces.

▶ **Definition 10** (Cones). A *cone* $(V, +, \cdot, 0)$ is a set V together with an associative and commutative operation $+$ with unit 0 and with a multiplication by real positive scalars $\cdot$ distributive over $+$. We have two more axioms: the *cancellation law* $\forall u, v, w \in V, v + u = w + u \Rightarrow v = w$ and the *strictness* $\forall v, w \in V, v + w = 0 \Rightarrow v = w = 0$.

Any cone C admits a natural partial order structure $\leq$ defined as follows: $u \leq v$ if and only if there exists w such that $v = u + w$. We will consider normed cones which are complete with respect to increasing sequences (*chains*) in this order which are of bounded norm.

▶ **Definition 11** (Normed cones, ω-complete). A *normed cone* C is a cone together with a function $\|-\| : C \to \mathbb{R}_+$ satisfying (i) $\forall v \in C, \|v\| = 0 \Leftrightarrow v = 0$; (ii) $\forall r \in \mathbb{R}_+, v \in C, \|r \cdot v\| = r\|v\|$; (iii) $\forall u, v \in C, \|u + v\| \leq \|u\| + \|v\|$; (iv) $u \leq v \Rightarrow \|u\| \leq \|v\|$. A cone is ω-*complete* if (i) for all chain $(u_n)_{n \in \mathbb{N}}$ such that $\{\|u_n\|\}_{n \in \mathbb{N}}$ is bounded, there exists a least upper bound (lub) $\bigvee_n u_n$ and (ii) $\|\bigvee_n u_n\| = \bigvee_n \|u_n\|$.

Note that the norm $\|-\|$ is ω-continuous. All the cones we are going to consider in the following are ω-complete and normed. ω-continuous linear maps form the natural notion of morphism between such structures. Note that linearity implies monotonicity in the natural order.

▶ **Definition 12** (ω-continuous linear maps). For C, D ω-complete normed cones, an ω-continuous linear map $f : C \to D$ is a linear map such that for every chain $(u_n)_{n \in \mathbb{N}}$ for which $\bigvee_n u_n$ exists, $\bigvee_n f(u_n)$ exists and is equal to $f(\bigvee_n u_n)$.

The dual of an ω-complete normed cone is defined in the usual way. We will admit the following result:

▶ **Proposition 13** (ω-complete dual)**.** *If C is an ω-complete normed cone, the cone of ω-continuous linear maps $\{\phi : C \to \mathbb{R}_+\}$ with the norm $\|\phi\| = \inf_v \{c \geq 0 \mid |\phi(v)| \leq c \, \|v\|\}$ is ω-complete.*

ω-complete normed cones and ω-continuous linear maps form a category denoted by $\omega\mathbf{CC}$. The dual operation gives rise to a contravariant endofunctor $-^* : \omega\mathbf{CC} \to \omega\mathbf{CC}$. If $f : C \to D$ is an $\omega\mathbf{CC}$ arrow, $f^* : D^* \to C^*$ is defined by $f^*(\phi) = \phi \circ f$. For all $\phi \in D^*$ and $x \in C$, one has

$$
\begin{aligned}
\|f^*(\phi)(x)\| &= \|(\phi \circ f)(x)\| \\
&\leq \|\phi\| \, \|f(x)\| \quad (\phi \ \omega\text{-continuous}) \\
&\leq \|\phi\| \, \|f\| \, \|x\| \quad (f \ \omega\text{-continuous})
\end{aligned}
$$

Therefore, $\|f^*\| \leq \|f\|$. We now introduce the cones we are going to work with in the remainder of the paper.

4.2   Cones of measures and of measurable functions

Let us fix a measure space (X, Σ, p) with p finite. Much of the constructions in the rest of the paper rely on dualities between the cones of measurable functions $L_1^+(X, \Sigma, p), L_\infty^+(X, \Sigma, p)$ and cones of measures $\mathcal{M}^{\ll p}(X, \Sigma), \mathcal{M}_{UB}^p(X, \Sigma)$. Let us introduce these cones in more detail.

Cones of measurable functions

Two positive measurable maps $f, f' : (X, \Sigma, p) \to \mathbb{R}_+$ are said to be *p-equivalent* if $p\{x \mid f(x) \neq f'(x)\} = 0$. Such as map $f : (X, \Sigma) \to \mathbb{R}$ is *p-integrable* if $\int_X f \, dp < \infty$. Clearly, being p-integrable is preserved by p-equivalence. The elements of the cone $L_1^+(X, \Sigma, p)$ are p-equivalence classes of real-valued integrable maps. $L_1^+(X, \Sigma, p)$ is normed by $\|f\|_1 = \int_X f \, dp$. The dominated convergence theorem implies that $L_1^+(X, \Sigma, p)$ is an ω-complete normed cone.

A positive measurable map f is *p-essentially bounded* if there exists $C \geq 0$ such that $p\{x \mid f(x) > C\} = 0$. The elements of the cone $L_\infty^+(X, \Sigma, p)$ are p-equivalence classes of real-valued essentially bounded maps. The norm is given by $\|f\|_\infty = \inf \{C \geq 0 \mid f(x) \leq C \ p\text{-a.s.}\}$.

Cones of measures

Closely related to the cones above are cones of absolutely continuous measures $\mathcal{M}^{\ll p}(X, \Sigma)$ and bounded measures $\mathcal{M}_{UB}^p(X, \Sigma)$. $\mathcal{M}^{\ll p}(X, \Sigma)$ is the cone of finite measures which are absolutely continuous with respect to p, with norm given by $\|q\|_\ll = q(X)$. $\mathcal{M}_{UB}^p(X, \Sigma)$ is the cone of finite measures which are uniformly bounded by a finite multiple of p, with norm given by $\|q\|_{UB} = \inf \{c \geq 0 \mid q \leq cp\}$. The ω-completeness of these cones will appear as a byproduct of the duality to be proved in the next.

4.3   Duality between L_1 and L_∞ cones

In the following, we will denote by $\mathbf{L}_1^+$ the full subcategory of $\omega\mathbf{CC}$ having as objects cones $L_1^+(X, \Sigma, p)$ and by $\mathbf{L}_\infty^+$ the full subcategory of $\omega\mathbf{CC}$ having as objects cones $L_\infty^+(X, \Sigma, p)$. As indicated before, the construction of the duality goes through cones of absolutely continuous measures $\mathcal{M}^{\ll p}(X, \Sigma)$ and bounded measures $\mathcal{M}_{UB}^p(X, \Sigma)$.

▶ **Theorem 14.** *$\mathcal{M}^{\ll p}(X, \Sigma)$ is isometrically isomorphic to $L_1^+(X, \Sigma, p)$ and $\mathcal{M}_{UB}^p(X, \Sigma)$ is isometrically isomorphic to $L_\infty^+(X, \Sigma, p)$.*

Proof. The Radon-Nikodym derivative induces a map $q \in \mathcal{M}^{\ll p}(X, \Sigma) \mapsto \frac{dq}{dp}$ which is linear, injective and norm-preserving, as is its inverse map $u \in L_1^+(X, \Sigma, p) \mapsto u \cdot p$. Therefore, $\mathcal{M}^{\ll p}(X, \Sigma)$ is isometrically isomorphic to $L_1^+(X, \Sigma, p)$. The isomorphism between $\mathcal{M}_{UB}^p(X, \Sigma)$ and $L_\infty^+(X, \Sigma, p)$ is proved similarly. We only show that $\frac{d-}{dp}$ is norm preserving. For $q \in \mathcal{M}_{UB}^p(X, \Sigma)$, we have $\left\| \frac{dq}{dp} \right\|_\infty = \inf \left\{ C \geq 0 \mid \frac{dq}{dp} \leq C \text{ } p\text{-a.s.} \right\}$. If $\frac{dq}{dp} \leq C$ then $q \leq C \cdot p$. By definition, the least such C is $\|q\|_{UB}$. ◀

As a consequence, both cones of finite measures are ω-complete. We are now in position to state the following variant of the Riesz representation theorem for ω-complete cones:

▶ **Theorem 15.** *We have the following isometric isomorphisms:* $L_1^{+,*}(X, \Sigma, p) \cong \mathcal{M}_{UB}^p(X, \Sigma)$ *and* $L_\infty^{+,*}(X, \Sigma, p) \cong \mathcal{M}^{\ll p}(X, \Sigma)$.

We will only prove the first isomorphism, the second one being proved similarly.

Proof. We construct an isometric isomorphism $\iota : L_1^{+,*}(X, \Sigma, p) \to \mathcal{M}_{UB}^p(X, \Sigma)$. Let us set $\iota(\phi) = B \mapsto \phi(\mathbb{1}_B)$. It is clearly linear. Let us show that $\iota(\phi)$ is countably additive. For $(B_n)_{n \in \mathbb{N}}$ a countable family of pairwise disjoint subsets, we have $\iota(\phi)(\cup_{n \in \mathbb{N}} B_n) = \phi(\mathbb{1}_{\cup_{n \in \mathbb{N}} B_n})$. Clearly, the family $\lambda_k = \sum_{i=1}^k \mathbb{1}_{B_n}$ is increasing, measurable and is bounded by $\mathbb{1}_X$. Therefore $\bigvee \lambda_k = \sum_n \mathbb{1}_{B_n}$ exists and by ω-continuity of ϕ, $\phi(\mathbb{1}_{\cup_{n \in \mathbb{N}} B_n}) = \sum_n \phi(\mathbb{1}_{B_n})$. ι is injective: let $v \in L_1^+(X, \Sigma, p)$ s.t. $\phi(v) \neq \phi'(v)$. v, being integrable can be approximated by an increasing sequence of simple functions. We deduce there must exists B such that $\phi(\mathbb{1}_B) \neq \phi'(\mathbb{1}_B)$. Let us prove that $\iota(\phi) \in \mathcal{M}_{UB}^p(X, \Sigma)$. A monotone convergence argument shows that for all $v \in L_1^+(X, \Sigma, p)$, we have the identity $\phi(v) = \int_X v \, d\iota(\phi) < \infty$. We also have $p(B) = 0 \Rightarrow \iota(\phi)(B) = 0$ hence $\iota(\phi) \ll p$. We deduce from the two previous facts that $\frac{d\iota(\phi)}{dp} \in L_\infty^+(X, \Sigma, q)$, which by Theorem 14 implies that $\iota(\phi) \in \mathcal{M}_{UB}^p(X, \Sigma)$. It remains to prove that ι is surjective: for $q \in \mathcal{M}_{UB}^p(X, \Sigma)$, we have trivially for $\phi_q = v \mapsto \int_X v \, dq$ that $\iota(\phi_q) = q$. ◀

An immediate consequence is that $L_1^{+,*}(X, \Sigma, p) \cong L_\infty^+(X, \Sigma, p)$ and reciprocally. This is notoriously false in the case of general Banach spaces. What makes everything work here is that we restrict to ω-continuous linear functionals.

Theorem 15 gives rise to a *pairing* between the spaces of integrable and bounded functions.

▶ **Definition 16** (Pairing). The *pairing* of $L_\infty^+(X, \Sigma, p)$ and $L_1^+(X, \Sigma, p)$ is a bilinear map $\langle \cdot, \cdot \rangle_X : L_\infty^+(X, \Sigma, p) \times L_1^+(X, \Sigma, p) \to \mathbb{R}$ given by $\langle u, v \rangle_X = \int_X u \cdot v \, dp$. It is continuous and ω-continuous in both arguments.

This pairing gives rise to a notion of *adjoint*:

▶ **Proposition 17** (Adjoints). *The duality functor* $-^* : \omega\mathbf{CC} \to \omega\mathbf{CC}$ *restricts to a contravariant functor* $-^\dagger : \mathbf{L}_1^+ \to \mathbf{L}_\infty^+$ *such that for all* $\mathbf{L}_1^+$ *arrow* $A : L_1^+(X, \Sigma, p) \to L_1^+(Y, \Lambda, q)$, $A^\dagger : L_\infty^+(Y, \Lambda, q) \to L_\infty^+(X, \Sigma, p)$ *is the unique adjoint arrow such that for all* $u \in L_1^+(X, \Sigma, p), v \in L_\infty^+(Y, \Lambda, q)$, $\langle v, Au \rangle_Y = \langle A^\dagger v, u \rangle_X$.

Conversely, for all $A : L_\infty^+(Y, \Lambda, q) \to L_\infty^+(X, \Sigma, p)$ *there is a unique adjoint* $^\dagger A : L_1^+(X, \Sigma, p) \to L_1^+(Y, \Lambda, q)$ *such that the equation above is verified.*

We will not prove Proposition 17 here but simply sketch how adjoints are constructed. Given $A : L_1^+(X, \Sigma, p) \to L_1^+(Y, \Lambda, q)$ and using that $L_\infty^+(X, \Sigma, p) \cong L_1^{+,*}(X, \Sigma, p)$ we set:

$$A^\dagger = v \in L_\infty^+(Y, \Lambda, q) \mapsto \left(u \in L_1^+(X, \Sigma, p) \mapsto \langle v, Au \rangle \right) \tag{4}$$

Note that $A^\dagger(v)$ is written in dual form. The adjoint of $^\dagger A : L_\infty^+(Y, \Lambda, q) \to L_\infty^+(X, \Sigma, p)$ is defined similarly. Finally, observe that setting Y to be the one point space in Proposition 17 implies Theorem 15.

4.4 Operator interpretations of kernels

In [3], it is shown that kernels that respect a condition of "non-singularity" correspond to particular linear operators between cones. This view on kernels allows to leverage the language of functional analysis to approximate and reason on these objects. We improve on this by showing that the functional interpretation of a non-singular kernel corresponds to functors defined on **Krn** and valued in $\mathbf{L}_1^+$ and $\mathbf{L}_\infty^+$. These new developments will be put to use in Section 5 where we will show that Bayesian inversion (Theorem 9) maps through this functorial correspondence to the adjunction of Proposition 17.

We introduce contravariant functors $\mathsf{T}_\infty : \mathbf{Krn} \to \mathbf{L}_\infty^+$ and $\mathsf{T}_1 : \mathbf{Krn} \to \mathbf{L}_1^+$ mapping kernels into norm 1 ω-continuous operators. T_∞ is defined on objects (X,p) by $\mathsf{T}_\infty(X,p) = L_\infty^+(X,p)$, while $\mathsf{T}_1(X,p) = L_1^+(X,p)$. For $f : (X,p) \rightarrowtail (Y,q)$ a **Krn** arrow, we set

$$\mathsf{T}_\infty(f) = v \in L_\infty^+(Y,q) \mapsto \left(x \mapsto \int_Y v\, df(x) \right). \tag{5}$$

Note that this is well-defined, as $f(x)$ is p-a.s. absolutely continuous with respect to q (Lemma 3). $\mathsf{T}_1(f)$ is defined similarly but acts on $L_1^+(Y,q)$.

▶ **Proposition 18.** T_∞ and T_1 are functors $\mathsf{T}_\infty : \mathbf{Krn} \to \mathbf{L}_\infty^+$ and $\mathsf{T}_1 : \mathbf{Krn} \to \mathbf{L}_1^+$ ranging in operators of norm 1.

Proof. We first check that T_∞ and T_1 are well-typed. Let us fix $f : (X,p) \rightarrowtail (Y,q)$ in **Krn**. $\mathsf{T}_\infty(f)$ and $\mathsf{T}_1(f)$ as defined above are clearly linear. We start with T_1. We have for all $v \in L_1^+(Y,q)$ that

$$\|\mathsf{T}_1(f)(v)\|_1 = \int_X \left(\int_Y v\, df(x) \right) dp = \int_Y v\, d(\mu_Y \circ \mathsf{G}(f))(p) = \int_Y v\, dq = \|v\|_1$$

Therefore $\|\mathsf{T}_1(f)\| = 1$. Prop. 5.2 of [3] ensures that $\mathsf{T}_1(f)$ is ω-continuous. Let us treat the case of T_∞. For all $v \in L_\infty^+(Y,q)$, the inequality $\int_Y v\, df(x) \le \|v\|_\infty$ is verified. This implies

$$\|\mathsf{T}_\infty(f)(v)\|_\infty = \inf \left\{ C \ge 0 \mid \int_Y v\, df(x) \le C\ p\text{-a.s.} \right\} \le \|v\|_\infty$$

therefore $\|\mathsf{T}_\infty(f)\| \le 1$. The upper bound is reached for $v = \mathbb{1}_Y$, therefore $\|\mathsf{T}_\infty(f)\| = 1$. ω-continuity of $\mathsf{T}_\infty(f)$ follows from the dominated convergence theorem. In the following, both T_1 and T_∞ are denoted by T. If we denote by $id' = \delta \circ id$ an identity in $\mathcal{K}\ell$, $\mathsf{T}(id')(v) = v$. Let $f : (X,p) \rightarrowtail (Y,q)$, $g : (Y,q) \rightarrowtail (Z,r)$ be two arrows in **Krn**. By definition of T and using naturality of μ, we get:

$$\mathsf{T}(g \circ f)(v)(x) = \int_Z v\, d(g \circ_M f(x)) = \int_{y \in Y} \left(\int_Z v\, dg(y) \right) df(x) = (\mathsf{T}(f)\mathsf{T}(g))(v)(x)$$

Therefore, T is a well-defined functor.   ◀

We will call operators in the range of T_∞ and T_1 *abstract Markov kernels*. Postcomposing T_∞ with $^\dagger-$ (or T_1 with $-^\dagger$) yields the covariant ("forward") interpretation of **Krn** arrows, called *Markov operators*. The restriction of T_∞ to $\mathbf{Krn}_\delta$ is familiar, as it reduces in this case to the precomposition functor $P_\infty : \mathbf{Krn}_\delta \to \mathbf{L}_\infty^+$ acting on $\mathbf{Krn}_\delta$ (hence deterministic) arrows $f : (X,p) \rightarrowtail_\delta (Y,q)$ by $P_\infty(f) = v \mapsto v \circ f$. Postcomposing P_∞ with $^\dagger-$ yields the conditional expectation functor $\mathbb{E}_1 : \mathbf{L}_1^+ \to \mathbf{L}_1^+$. Similarly, functors P_1 and $\mathbb{E}_\infty$ can are constructed by considering T_1 instead of T_∞. We invite the reader to compare our developments with

Section 3 and 4 of [3] for more details. We sum up the developments so far in the following (non-commuting) diagram:

$$(6)$$

We conclude this section by indicating that the duality between $\mathbf{L}_1^+$ and $\mathbf{L}_\infty^+$ generalises to arbitrary dual pairs $\mathbf{L}_p^+, \mathbf{L}_q^+$ [2]. We conjecture that the functors $\mathsf{T}_1, \mathsf{T}_\infty$ have counterparts in this more general setting. In the next section, we give a functional interpretation of Bayesian inversion through these functors.

5 Bayesian inversion as duality

As shown in Section 3, Bayesian inversion is a symmetrised disintegration, which by Equation 2 corresponds to a measurable family of conditional probabilities. As these are a fundamental tool of the modern probabilistic toolkit, a natural question is to find a corresponding process in the functional analytic setting of norm-1 operators between the ω-complete cones L_1^+ and L_∞^+. It is well-known that conditional expectation can be framed as a projection operator (e.g. in the L_2 case, see [10]). The result we are about to prove provides a fresh perspective on this classical problem: the Bayesian inverse of a kernel corresponds to the adjoint of its functional form.

▶ **Theorem 19** (Inversion as duality). *Let X, Y be standard Borel spaces and $f : (X, p) \to (Y, q)$ a* **Krn** *arrow. We have:*

$$^\dagger\mathsf{T}_\infty(f) = \mathsf{T}_1(f^\dagger)$$

I.e. $\mathsf{T}_\infty(f)$ *is adjoint to* $\mathsf{T}_1(f^\dagger)$.

Proof. Let us recall the types of the objects:

$$
\begin{aligned}
\mathsf{T}_\infty(f) &: L_\infty^+(Y, q) \to L_\infty^+(X, p) \\
^\dagger\mathsf{T}_\infty(f) &: L_1^+(X, p) \to L_1^+(Y, q)
\end{aligned}
$$

It is enough to prove that for all $u \in L_1^+(Y, q)$, for all $v \in L_\infty^+(X, p)$,

$$\langle v, \mathsf{T}_\infty(f)(u) \rangle_X = \langle \mathsf{T}_1(f^\dagger)(v), u \rangle_Y$$

Unfolding, we must prove:

$$
\begin{aligned}
& \int_X v \mathsf{T}_\infty(f)(u) dp & = & \int_Y \mathsf{T}_1(f^\dagger)(v) u dq \\
\Leftrightarrow \quad & \int_{x \in X} v(x) \left(\int_Y u df(x) \right) dp & = & \int_{y \in Y} \left(\int_X v \, df^\dagger(y) \right) u(y) dq \\
\Leftrightarrow \quad & \int_{x \in X} \left(\int_{y \in Y} v(x) u(y) df(x) \right) dp & = & \int_{y \in Y} \left(\int_{x \in X} v(x) u(y) df^\dagger(y) \right) dq
\end{aligned}
$$

Let $\gamma \in \Gamma(p, q)$ be the coupling corresponding to $f : (X, p) \to (Y, q)$. Continuing the string of equivalences above, we must prove:

$$\int_{x \in X} \left(\int_{y \in Y} v(x) u(y) df(x) \right) d\mathsf{G}(\pi_X)(\gamma) \quad = \quad \int_{y \in Y} \left(\int_{x \in X} v(x) u(y) df^\dagger(y) \right) d\mathsf{G}(\pi_Y)(\gamma)$$

Now, recall that by Equation 1, the disintegration theorem gives us for all integrable $e : (X \times Y, \gamma) \to \mathbb{R}$ that

$$
\begin{aligned}
\int_{X \times Y} e\, d\gamma
&= \int_{x \in X} \left(\int_{\pi_X^{-1}(x)} e\, df(x) \right) d\mathsf{G}(\pi_X)(\gamma) && \text{(disintegrating } \gamma \text{ along } \pi_X) \\
&\overset{1}{=} \int_{x \in X} \left(\int_Y e(x, \_)\, df(x) \right) d\mathsf{G}(\pi_X)(\gamma) && \text{(disintegrations live on the fiber)} \\
&= \int_{y \in Y} \left(\int_{\pi_Y^{-1}(y)} e\, df^\dagger(y) \right) d\mathsf{G}(\pi_Y)(\gamma) && \text{(disintegrating } \gamma \text{ along } \pi_Y) \\
&\overset{2}{=} \int_{y \in Y} \left(\int_X e(\_, y)\, df^\dagger(y) \right) d\mathsf{G}(\pi_Y)(\gamma) && \text{(disintegrations live on the fiber)}
\end{aligned}
$$

Taking $e(x, y) = v(x)u(y)$ and using equations marked 1 and 2 above concludes the proof. ◀

Some comments are in order. Note that the disintegration and Bayesian inversion theorems rely on some strong assumptions on the underlying spaces–here, we assume the spaces to be standard Borel; Culbertson & Sturtz [4] work in the setting of perfect measure spaces and equiperfect kernels. However, the cone duality works for any measure space! We conjecture that these regularity conditions are necessary if one wishes to extract a measurable kernel from a Markov operator or dually from an abstract Markov kernel.

6 Conclusion

We have established that the functional representation of measurable kernels as operators acting on ω-complete cones presented in [3] is functorial. Two variants of the functor exist, mapping kernels to operators acting either on bounded functions or on integrable ones. The category of 'typed' kernels on which these functors are defined allows to state elegantly the famous disintegration theorem and its generalisation, Bayesian inversion. What's more, we uncovered the categorical underpinnings of Bayesian inversion as particular natural transformations mapping kernels to couplings and reciprocally. Finally, we have shown that Bayesian inversion amounts in the functional world to adjunction.

Several further developments suggest themselves. First of all, It remains to be seen whether our construction generalises from the duality L_1^+/L_∞^+ to arbitrary pair of dual cones L_p^+/L_q^+ (e.g. the pair $p = q = 2$ which allows one to talk about reversible kernels), and can prove a stronger statement, namely $\mathsf{T}_p(f^\dagger) = \mathsf{T}_q(f)^\dagger$ for *all* conjugate exponents p, q. Another line of thought is to connect these results with some of the authors recent's work [7, 5]. In particular, instantiating the kernel-theoretic framework with the Dirichlet process [7], might provide insight into the operator-theoretic counterpart of so-called nonparametric methods in Bayesian learning. On a different note, this process has the type of a *natural* and *continuous* kernel and admits a convenient finitary characterisation. We are eager to study how these properties map through the operator interpretation.

 ── **References** ──

1   C. Aliprantis and K. Border. *Infinite dimensional analysis*, volume 32006. Springer, 1999.
2   C. Badescu and P. Panangaden. Abstract Markov processes with L_p spaces. Private communication, 2012.
3   P. Chaput, V. Danos, P. Panangaden, and G. Plotkin. Approximating Markov Processes by averaging. *Journal of the ACM*, 61(1), January 2014. 45 pages.
4   J. Culbertson and K. Sturtz. A categorical foundation for Bayesian probability. *Applied Categorical Structures*, pages 1–16, 2012.
5   F. Dahlqvist, V. Danos, and I. Garnier. Giry and the machine. Feb 2016. To appear in the proceedings of MFPS XXXII.

6   V. Danos, J. Desharnais, F. Laviolette, and P. Panangaden. Bisimulation and cocongruence
       for probabilistic systems. *Information and Computation*, 204(4):503–523, 2006.

7   V. Danos and I. Garnier. Dirichlet is natural. *Electronic Notes in Theoretical Computer
       Science*, 319:137 – 164, 2015. MFPS XXXI.

8   C. Dellacherie and P.A. Meyer. *Probabilities and Potential, C: Potential Theory for Discrete
       and Continuous Semigroups*. North-Holland Mathematics Studies. Elsevier Science, 2011.

9   M. Giry. A categorical approach to probability theory. In *Categorical Aspects of Topology
       and Analysis*, number 915 in Lecture Notes In Math., pages 68–85. Springer-Verlag, 1981.

10 O. Kallenberg. *Foundations of Modern Probability*. Springer, 1997.

11 A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Text in Mathematics*.
       Springer, 1995.

12 D. Kozen. Semantics of probabilistic programs. *Journal of Computer and System Sciences*,
       22(3):328–350, 1981.

13 S. Mac Lane. *Categories for the Working Mathematician*, volume 111 of *Graduate Texts in
       Mathematics*. Springer, 1998.

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