On Disintegration in Probabilistic Semantics

Dexter Kozen

Computer Science Department
Cornell University
Ithaca, New York 14853-7501, USA

Abstract
We give a version of disintegration based on weaker assumptions than previously known. The result is derived using an enhanced version of the Radon-Nikodým theorem that makes explicit the approximating functions, thereby allowing the coordinated construction of multiple simultaneous derivatives. Using this, a category of joint distributions can be independently defined in which a category of Markov kernels on countably generated spaces is fully and faithfully embedded.

Keywords: probabilistic programming, disintegration, Bayesian inference, conditioning

1 Introduction

Bayesian inference and conditioning are important tools in probabilistic programming. Modern probabilistic languages for machine learning, e.g. Church [7] and Anglican [12], generally incorporate these tools in some form. To formalize the semantics of such languages, several authors have proposed categories for modeling Bayesian inference and conditioning via disintegration.

One such category is Kn, recently proposed by Dahlqvist et al. [5]. The objects of Kn are probability spaces (X, A, µ) with A a standard Borel space. The morphisms (X, A, µ) → (Y, B, ν) are equivalence classes modulo µ-nullsets of Markov kernels P : X → Y such that

\[ ν(B) = \int_{s \in X} P(s, B) \mu(ds). \]

Dahlqvist et al. [5] use disintegration to show that Kn is a dagger category with involutive functor ⨆ : Kn → Knop. Thus P† : Y → X is a kernel in the opposite direction that models Bayesian inference of an input distribution conditioned on output samples. This works even for continuous distributions in which outcomes can occur with zero probability.

The kernel P gives rise to a joint distribution JP on X × Y with marginals µ and ν in a natural way. The disintegration theorem says that under certain assumptions—usually that the spaces are standard Borel, as in e.g. [1, 2, 5, 6, 9, 10]—this construction can be inverted: for any joint distribution θ on X × Y with marginals µ and ν, there is a kernel P, unique up to a µ-nullset, such that θ = JP. We show here that the inverse construction can be carried out under a weaker assumption than previously known, namely that the σ-algebra

1 Thanks to Fredrik Dahlqvist, Vincent Danos, Nate Foster, Justin Hsu, Bart Jacobs, Bobby Kleinberg, Radu Mardare, Prakash Panangaden, Daniel Roy, Alexandra Silva, and Steffen Smolka. Thanks to the Bellairs Research Institute of McGill University for providing a wonderful research environment.

2 Email: kozen@cs.cornell.edu

3 Research supported by NSF grant CCF1637532.
$\mathcal{B}$ of measurable sets of $Y$ is countably generated; equivalently, $\mathcal{B}$ is the Baire space of a countably generated topology.

More accurately, we show that $\text{Krn}$—or rather, its extension to all countably generated spaces—is fully and faithfully embedded in a category of joint distributions $\text{JD} \text{ist}$ that can be defined independently. The objects of $\text{JD} \text{ist}$ are measure spaces $(X, \mathcal{A}, \mu)$ and the morphisms $(X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ are joint distributions $\theta$ on $X \times Y$ with marginals $\mu$ and $\nu$. The category $\text{JD} \text{ist}$ is a symmetric monoidal category with product distributions as tensors and transpose as symmetry: $\theta'(A \times B) = \theta(B \times A)$. Thus $\mathbf{J}$ is an isomorphism of categories between $\text{Krn}$ and the full subcategory of $\text{JD} \text{ist}$ on countably generated spaces.

This construction was previously known for subcategories of $\text{JD} \text{ist}$ for spaces satisfying various stronger restrictions [1,3,4,6,10]. The most common assumption is that the spaces are standard Borel. The most general result of this type is that of Culbertson and Sturtz [4], who assume countably generated spaces but that the measures are perfect. Abramsky et al. [1] studied a subcategory of $\text{JD} \text{ist}$ on Polish spaces called $\text{PRel}$, but used disintegration to define composition. We show that composition in $\text{JD} \text{ist}$ can be defined independently, without reference to disintegration and without any restriction on the spaces whatsoever. The definition uses an enhanced Radon-Nikodým construction, which we describe in §3.1.

The main technical challenge is to coordinate the simultaneous creation of multiple Radon-Nikodým derivatives. For each fixed $D \in \mathcal{B}$, one can create a suitable measurable function $P(\cdot, D)$ using the standard derivative construction as given for example in [8, 11]. But to get a Markov kernel, one must also coordinate these constructions for different $D$ to make $P(s, \cdot)$ countably additive. This requires an enhancement of the Radon-Nikodým theorem to include information that is normally available in the proof but lost in the statement of the theorem. This is where the assumption that the space is countably generated is used. We state and reprove this enhanced version in §3.1.

2 Notation

Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. We abbreviate $(X, \mathcal{A})$ by $X$ when $\mathcal{A}$ is understood. The letters $A, B, C, D$ denote measurable sets. The space $(X, \mathcal{A})$ is a standard Borel space if $\mathcal{A}$ are the Borel sets of a Polish space (separable and completely metrizable). The space $(X, \mathcal{A})$ is countably generated if there exists a countable set $\mathcal{A}_0 \subseteq \mathcal{A}$ such that $\mathcal{A}$ is the smallest $\sigma$-algebra containing $\mathcal{A}_0$. All standard Borel spaces are countably generated.

A Markov kernel $P : X \to Y$ is a map $P : X \times \mathcal{B} \to [0,1]$ such that

1. for fixed $s \in X$, $P(s, \cdot)$ is a probability measure on $Y$,
2. for fixed $B \in \mathcal{B}$, $P(\cdot, B)$ is a measurable function on $X$.

These properties allow kernels to be sequentially composed by Lebesgue integration. The measurable spaces and Markov kernels form a category $\text{SRel}$ [6,10], which is isomorphic to the Kleisli category of the Giry monad. We write $P : X \to Y$ for the kernel $P$ regarded as a morphism in this category.

For $P : X \to Y$ a Markov kernel and $\mu$ a finite measure on $X$, write $\mu \circ P$ for the measure on $Y$ such that

$$\mu \circ P(B) = \int_X P(s, B) \mu(ds).$$

This gives a bounded linear map $(- \circ P) : \text{Meas} X \to \text{Meas} Y$ that is monotone and continuous in both the metric and Scott topologies.

For $A \in \mathcal{A}$, let $A$ also denote the subidentity kernel

$$A(s, B) = 1_X(s, A \cap B) = \begin{cases} 1, & s \in A \cap B, \\ 0, & s \notin A \cap B. \end{cases}$$

Then for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$,

$$\mu \circ A \circ P(B) = \int_A P(s, B) \mu(ds). \quad (1)$$

The category $\text{JD} \text{ist}$ is the category whose objects are probability spaces $(X, \mathcal{A}, \mu)$ and whose morphisms $(X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ are joint distributions or couplings on $X \times Y$ with marginals $\mu$ and $\nu$. We will define composition in $\text{JD} \text{ist}$ formally in §5, but we can already define the embedding functor $\mathbf{J} : \text{Krn} \to \text{JD} \text{ist}$. It is the identity on objects, and for morphisms $P : X \to Y$, let $P' : X \to X \times Y$ be the kernel that behaves like
Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces, $\mathcal{B}$ countably generated. Let $P, Q : X \to Y$ be Markov kernels and $\mu$ a probability measure on $X$. The following are equivalent:

(i) $P \equiv_{\mu} Q$

(ii) $P \equiv_{\mu} Q$ for all $\xi$ absolutely continuous with respect to $\mu$ (notation: $\xi \ll \mu$)

(iii) for all $A \in \mathcal{A}, \mu ; A ; P = \mu ; A ; Q$

(iv) $JP = JQ$, considering $P$ and $Q$ as Kron-morphisms $(X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$, where $\nu = \mu ; P = \mu ; Q$.

**Proof.** The equivalence of (i) and (ii) is clear from the definition (3).

For (i) $\implies$ (iii), suppose that $P \equiv_{\mu} Q$. Let $E = \{s \mid P(s) = Q(s)\}$. By definition, $\mu(X \setminus E) = 0$. For all $A \in \mathcal{A}$ and $B \in \mathcal{B},$

$$(\mu ; A ; P)(B) = \int_A P(s, B) \mu(ds) = \int_{A \cap E} P(s, B) \mu(ds) + \int_{A \setminus E} P(s, B) \mu(ds) = \int_{A \cap E} Q(s, B) \mu(ds) + \int_{A \setminus E} Q(s, B) \mu(ds) = \int_A Q(s, B) \mu(ds) = (\mu ; A ; Q)(B).$$

The left-hand terms of (4) and (5) agree because the integral is restricted to $E$, and the right-hand terms are 0 because $A \setminus E$ is a $\mu$-nullset. As $B$ was arbitrary, $\mu ; A ; P = \mu ; A ; Q$.

Conversely, for (iii) $\implies$ (i), if $P \not\equiv_{\mu} Q$, then $\mu(\{s \mid |P(s, B) - Q(s, B)| \geq 1/n\}) > 0$ for some $B \in \mathcal{B}_0$ and $n > 0$. Letting $A$ be this set, we have

$$\int_A |P(s, B) - Q(s, B)| \mu(ds) \geq \frac{1}{n} \mu(A) > 0,$$

so

$$(\mu ; A ; P)(B) = \int_A P(s, B) \mu(ds) \not\equiv \int_A Q(s, B) \mu(ds) = (\mu ; A ; Q)(B).$$

The equivalence of (iii) and (iv) follows from (2) and (1). \qed

Let $(X, \mathcal{A})$ be a measurable space. The countable measurable partitions $\mathcal{D}$ of $X$ form an upper semilattice ordered by refinement, denoted $\sqsubseteq$, with least common refinement as join, denoted $\sqcup$. We will often consider
the limiting behavior of functions defined on increasingly finer such partitions. For any such map \( \varphi \) taking values in a topological space, if \( \varphi(D_n) \) converges to the same value for all chains \( D_0 \subseteq D_1 \subseteq \cdots \) that eventually become sufficiently fine, we write \( \lim_{D} \varphi(D) \) for this value.

3 Disintegration

Recall from §2 that, for a \textbf{Krn}-morphism \( P : (X, A, \mu) \to (Y, B, \nu) \), \( J P \) is the joint measure on \( X \times Y \) whose value on measurable rectangles \( A \times B \) is

\[
J P(A \times B) = \int_A P(s, B) \mu(ds)
\]

with marginals \( \mu \) and \( \nu \); \( P \). Any measure \( \mu \) on \( X \) and kernel \( P : X \to Y \) give rise to a joint measure on \( X \times Y \) in this way.

Our goal in this section is to prove a version of the disintegration theorem that says that this construction is invertible under the assumption that \( B \) is countably generated.

**Theorem 3.1 (Disintegration)** Let \( (X, A) \) and \( (Y, B) \) be measurable spaces, \( B \) countably generated. Let \( \theta \) be a joint probability measure on \( X \times Y \) with left marginal \( \mu = \theta(\cdot \times Y) \). There exists a Markov kernel \( P : X \to Y \) such that

\[
\theta(A \times B) = \int_A P(s, B) \mu(ds).
\]

The kernel \( P \) is unique modulo \( \equiv \mu \).

The traditional formulation of disintegration, as presented for example in [2, 5, 9], is somewhat different from the formulation above. It states that for every measurable function \( f : X \to Y \) and probability measure \( \mu \) on \( X \), there exists a kernel \( Q : Y \to X \) that is a right inverse of \( f \) and is unique up to a nullset. In our treatment, the traditional version follows as a corollary (Corollary 4.1).

3.1 Radon-Nikodým Revisited

In order to invert \( J \), we need to coordinate the creation of multiple Radon-Nikodým derivatives. For each fixed \( D \in B \), one can create a measurable function \( P(-, D) \) satisfying (7) using the standard derivative construction (see e.g. [8]). But one must also coordinate these constructions for different \( D \) in order to make the resulting \( P(s, -) \) countably additive, so that we end up with a Markov kernel. The standard Radon-Nikodým theorem provides no hook for this. There is a stronger version proved in [11, Theorem 3, p. 258] that includes the Lebesgue decomposition theorem as an ancillary result. This version comes closer to our needs, but unfortunately we still require some information available in the proof that is lost in the statement of the theorem. We must therefore restate it here in a somewhat stronger form (Theorem 3.3 below).

Let \( \nu \) and \( \mu \) be finite measures on \( (X, A) \). For any \( B \in A \), consider the set

\[
\left\{ \frac{\nu(C)}{\mu(C)} \mid C \subseteq B, \ \mu(C) > 0 \right\} \subseteq \mathbb{R}.
\]

(8)

This set is nonempty iff \( \mu(B) > 0 \). In that case, the set has a finite infimum, since \( \nu(B)/\mu(B) \) is a member, but it may be unbounded above.

**Lemma 3.2** Let \( \nu \) and \( \mu \) be finite measures on \( (X, A) \). For any \( \varepsilon > 0 \), there exists a countable measurable partition \( D \) of \( X \) such that for all \( B \in D \) with \( \mu(B) > 0 \), the set (8) is bounded above and

\[
\sup_{\substack{C \subseteq B \\mu(C) > 0}} \frac{\nu(C)}{\mu(C)} - \inf_{\substack{C \subseteq B \\mu(C) > 0}} \frac{\nu(C)}{\mu(C)} \leq \varepsilon
\]

(9)

\[
\left( \sup_{\substack{C \subseteq B \\mu(C) > 0}} \frac{\nu(C)}{\mu(C)} - \inf_{\substack{C \subseteq B \\mu(C) > 0}} \frac{\nu(C)}{\mu(C)} \right) \left( \sup_{\substack{C \subseteq B \\mu(C) > 0}} \frac{\nu(C)}{\mu(C)} \right) \leq \varepsilon^2.
\]

(10)

Moreover, these properties are preserved under refinement; that is, if \( D \subseteq D' \) and \( D \) satisfies (9) and (10) for all \( B \in D \) with \( \mu(B) > 0 \), then the same is true of \( D' \).
Theorem 3.3 (Lebesgue-Radon-Nikodým) For $k \geq 1$, consider the signed measure $\nu - (\varepsilon \ln k) \mu$. By the Hahn decomposition theorem, there exist measurable partitions \( \{A_{i}^{+}, A_{i}^{-}\} \) of \( X \) such that $\nu - (\varepsilon \ln k) \mu$ is purely nonnegative on $A_{i}^{+}$ and purely nonpositive on $A_{i}^{-}$; that is, for all $C \subseteq A_{i}^{+}$, $\nu \geq (\varepsilon \ln k) \mu$ and for all $C \subseteq A_{i}^{-}$, $\nu \leq (\varepsilon \ln k) \mu$. We can assume $A_{i}^{+} = X$. Let the partition $D$ consist of the sets $\bigcap_{i=1}^{k} A_{i}^{+} \cap A_{i+1}^{-}$, $k \geq 0$, along with $\bigcap_{i=1}^{\infty} A_{i}^{+}$, which is a $\mu$-nullset. For any measurable $C \subseteq \bigcap_{i=1}^{k} A_{i}^{+} \cap A_{i+1}^{-}$, we have $(\varepsilon \ln k) \mu(C) \leq \nu(C) \leq (\varepsilon \ln(k+1)) \mu(C)$, so if $\mu(C) > 0$, then $\nu(C)/\mu(C)$ exists and lies in the interval $[\varepsilon \ln k, \varepsilon \ln(k+1)]$. Since $\ln(1+x) \leq x$ for all $x \geq 1$, the left-hand sides of (9) and (10) are bounded, respectively, by

$$\varepsilon \ln(k+1) - \varepsilon \ln k = \varepsilon \ln\left(\frac{k+1}{k}\right) = \varepsilon \ln(1 + \frac{1}{k}) \leq \varepsilon \frac{\varepsilon}{k} \leq \varepsilon$$

$$(\varepsilon \ln(k+1) - \varepsilon \ln k) (\varepsilon \ln(k+1)) = \varepsilon \ln(1 + \frac{1}{k}) \varepsilon \ln(k+1) \leq \varepsilon \frac{\varepsilon}{k} \cdot \varepsilon = \varepsilon^{2}.$$

Let $D$ be a countable measurable partition of $A$. For $s \in X$, define

$$F_{D} = \bigcup \{B \in D \mid \mu(B) > 0\}$$

$$(\frac{d\nu}{d\mu})_{D}^{+} (s) = \sum_{B \in D} \sup_{\mu(B) > 0, \mu(C) > 0} \frac{\nu(C)}{\mu(C)} \cdot 1_{X}(s, B)$$

$$(\frac{d\nu}{d\mu})_{D}^{-} (s) = \sum_{B \in D} \inf_{\mu(B) > 0, \mu(C) > 0} \frac{\nu(C)}{\mu(C)} \cdot 1_{X}(s, B).$$

The set $F_{D}$ is measurable with $\mu(X \setminus F_{D}) = 0$, and the functions $(d\nu/d\mu)^{+}_{D}$ and $(d\nu/d\mu)^{-}_{D}$ are measurable step functions that vanish outside $F_{D}$. From Lemma 3.2, we have that for sufficiently fine $D$,

$$(\frac{d\nu}{d\mu})_{D}^{+} - (\frac{d\nu}{d\mu})_{D}^{-} \leq \varepsilon$$

$$\left( (\frac{d\nu}{d\mu})_{D}^{+} - (\frac{d\nu}{d\mu})_{D}^{-} \right) \leq \varepsilon^{2}.$$

**Theorem 3.3 (Lebesgue-Radon-Nikodým)** There exist measures $\nu_{0}, \nu_{1}$, a measurable set $F \in A$, a measurable real-valued function $f$ defined on $X$, and a countable $\subseteq$-chain $D_{0} \subseteq D_{1} \subseteq \cdots$ such that

(i) (Lebesgue decomposition) $\nu_{0}$ and $\nu_{1}$ form a Lebesgue decomposition of $\nu$ on $F$ with respect to $\mu$; that is,

$$\nu = \nu_{0} + \nu_{1} \quad \nu_{0} \ll \mu \quad \nu_{1}(F) = 0 \quad \mu(X \setminus F) = 0;$$

(ii) (Radon-Nikodým theorem) $f(s) = 0$ for all $s \not\in F$ and

$$\int_{A} f(s) \mu(ds) = \nu_{0}(A), \ A \in A;$$

morever, the value of the integral is independent of the choice of the chain $D_{n}$, provided the $D_{n}$ are sufficiently fine;

(iii) (Uniform approximation) The sequence $(d\nu/d\mu)^{+}_{n} = (d\nu/d\mu)_{D_{n}}^{+}$ is monotone nondecreasing on $F$ and $(d\nu/d\mu)^{-}_{n} = (d\nu/d\mu)_{D_{n}}^{-}$ is monotone nonincreasing everywhere, and both sequences converge pointwise to $f$ and converge uniformly on $F$.

If $\nu \ll \mu$, we can take $\nu_{0} = \nu$ and $\nu_{1} = 0$ in (i), in which case (ii) gives

$$\int_{A} f(s) \mu(ds) = \nu(A).$$

---

4 The signed measure $\nu - \varepsilon k \mu$ already implies (9), which suffices for most purposes, and standard treatments use this. But $\nu - (\varepsilon \ln k) \mu$ gives the additional conclusion (10), which will be useful later in the definition of composition in JDist.
In this case, $f$ is the standard Radon-Nikodým derivative $d\nu/d\mu$.

The version [11, Theorem 3, p. 258] asserts (i) and (ii) only without reference to the approximants $(d\nu/d\mu)^{+}$ and $(d\nu/d\mu)^{-}$, but (iii) is essentially how it is proved. In fact, [11] gives three proofs. We give a fourth here for completeness.

A subtle point: The value of the integral (14) in (ii) is independent of the choice of the $\subseteq$-chain $D_n$ in (iii), but $F$ and $f$ are not, and there is no one choice that works uniformly for all $\subseteq$-chains. This is why $d\nu/d\mu$ is only defined up to a $\mu$-nullset. However, by taking least common refinements, one can find $F$ and $f$ that work for multiple $\subseteq$-chains. That is the reason for making (iii) explicit.

**Proof.** Let $(d\nu/d\mu)^{+}$ and $(d\nu/d\mu)^{-}$ be defined as in (12) and (13). By definition,

$$
\left( \frac{d\nu}{d\mu} \right)^{-}_{D}(s) \leq \sum_{B \in D} \frac{\nu(B)}{\mu(B)} \cdot 1_{X}(s, B) \leq \left( \frac{d\nu}{d\mu} \right)^{+}_{D}(s).
$$

If $D \subseteq D'$, then $F_{D'} \subseteq F_{D}$, $(d\nu/d\mu)^{-}_{D} \leq (d\nu/d\mu)^{+}_{D'}$, pointwise on $F_{D'}$, and $(d\nu/d\mu)^{-}_{D} \leq (d\nu/d\mu)^{+}_{D} \leq (d\nu/d\mu)^{-}_{D}$ pointwise everywhere. Moreover, by Lemma 3.2, for all $\varepsilon > 0$ there exists a sufficiently fine $D$ that $(d\nu/d\mu)^{+}_{D} - (d\nu/d\mu)^{-}_{D} < \varepsilon$ pointwise. It follows that for any countable chain $D_0 \subseteq D_1 \subseteq \cdots$ of sufficiently fine countable measurable partitions, we have that $(d\nu/d\mu)^{+}_{D}$ and $(d\nu/d\mu)^{-}_{D}$ converge pointwise to a measurable function $f = \inf_{n}(d\nu/d\mu)^{+}_{n}$ and converge uniformly on $F = \bigcap_{n} F_{D_n}$. In addition, the region of uniform convergence $F$ is of full $\mu$-measure, as $\mu(F) = \inf_{n} \mu(F_{D_n}) = \mu(X)$, and $f$ vanishes outside $F$. This establishes (iii).

For (i), if $\mu(C) = 0$, then $\nu(C \cap A_{k}^{c}) \leq (\varepsilon \ln k)\mu(C \cap A_{k}^{c}) = 0$ for all $k$, where the $A_{k}^{c}$ are the Hahn decomposition sets constructed in the proof of Lemma 3.2. Assuming $D$ refines one of the partitions defined in that lemma, we have $F \subseteq F_{D'} \subseteq \bigcup_{k} A_{k}^{c}$, so $\nu(C \cap F) = 0$. Taking $\nu_{0}(C) = \nu(C \cap F)$ and $\nu_{1}(C) = \nu(C \setminus F)$ give a Lebesgue decomposition of $\nu$ on $F$ with respect to $\mu$; in particular,

$$
\mu(C) = 0 \Rightarrow \nu(C) = \nu(C \setminus F) = \nu_{1}(C).
$$

For (ii), observe that for sufficiently fine $D$,

$$
\int_{A} \left( \frac{d\nu}{d\mu} \right)^{+}(s) \mu(ds) = \sum_{B \in D} \sup_{C \subseteq B, \mu(B) > 0, \mu(C) > 0} \frac{\nu(C)}{\mu(C)} \cdot \mu(A \cap B) \leq \sum_{B \in D} \left( \inf_{C \subseteq B, \mu(C) > 0} \frac{\nu(C)}{\mu(C)} + \varepsilon \right) \cdot \mu(A \cap B)
$$

$$
\leq \sum_{B \in D} \left( \frac{\nu(A \cap B)}{\mu(A \cap B)} + \varepsilon \right) \cdot \mu(A \cap B) \leq \nu(A) + \varepsilon \mu(A),
$$

thus the integral exists and is finite. Moreover,

$$
\int_{A} \left( \frac{d\nu}{d\mu} \right)^{-}(s) \mu(ds) = \sum_{B \in D_n} \sup_{C \subseteq B, \mu(B) > 0, \mu(C) > 0} \frac{\nu(C)}{\mu(C)} \cdot \mu(A \cap B) = \sum_{B \in D_n} \sup_{C \subseteq B, \mu(A \cap B) > 0} \frac{\nu(C)}{\mu(C)} \cdot \mu(A \cap B). \tag{16}
$$

The right-hand equality in (16) follows from the observation that all summands corresponding to $B \in D_n$ with $\mu(A \cap B) = 0$ vanish, whereas for those with $\mu(A \cap B) > 0$, the test $\mu(B) > 0$ is redundant. Specializing the supremum in (16) at $C = A \cap B$,

$$
\int_{A} \left( \frac{d\nu}{d\mu} \right)^{+}(s) \mu(ds) \geq \sum_{B \in D_n, \mu(A \cap B) > 0} \nu(A \cap B) = \nu(A) - \sum_{B \in D_n, \mu(A \cap B) = 0} \nu(A \cap B) \tag{15}
$$

$$
= \nu(A) - \sum_{B \in D_n, \mu(A \cap B) = 0} \nu_{1}(A \cap B) \geq \nu(A) - \sum_{B \in D_n} \nu_{1}(A \cap B) = \nu(A) - \nu_{1}(A) = \nu_{0}(A).
$$

6
Similarly,
\[
\int_A \left( \frac{\partial}{\partial p} \right)_n^-(s) \mu(ds) = \sum_{B \in D_n} \inf_{C \subseteq B, \mu(C) > 0} \frac{\nu(C)}{\mu(C)} \cdot \mu(A \cap B) = \sum_{B \in D_n} \inf_{\mu(A \cap B \cap F) > 0} \frac{\nu(C)}{\mu(C)} \cdot \mu(A \cap B \cap F).
\]

The former equality follows from an argument similar to (16), the latter from the fact that \( \mu \) vanishes outside \( F \). Specializing the infimum at \( C = A \cap B \cap F \), we have
\[
\int_A \left( \frac{\partial}{\partial p} \right)_n^-(s) \mu(ds) \leq \sum_{B \in D_n} \nu(A \cap B \cap F) \leq \sum_{B \in D_n} \nu_0(A \cap B) = \nu_0(A).
\]

Thus \( \nu_0(A) \) is the limit (14). \( \square \)

### 3.2 Proof of the Disintegration Theorem

Suppose we are given a joint distribution \( \theta \) on \( X \times Y \) with left marginal \( \mu = \theta(- \times Y) \). For \( D \in \mathcal{B} \) and countable measurable partition \( \mathcal{D} \) of \( X \), let
\[
P^+_n(-, D) = \left( \frac{d\theta(- \times Y)}{d\mu} \right)_D^+ = \left( \frac{d\theta(- \times Y)}{d\theta(- \times Y)} \mu \right)_D^+
\]
as defined in (12). By Theorem 3.3, a sequence \( D_0 \subseteq D_1 \subseteq \cdots \) can be chosen so that the \( P^+_n(-, D) \) converge pointwise uniformly to a measurable function \( P(-, D) \) on a set \( F_D \) of full \( \mu \)-measure. The set \( F_D \) depends on the choice of \( D_n \). A different sequence \( D'_n \) will give a different \( F'_D \); but the sequence \( D_n \cup D'_n \) will converge to \( P(-, D) \) on the smaller domain \( F_D \cap F'_D \), and \( F_D \cap F'_D \) is still of full \( \mu \)-measure.

Assuming that \( \mathcal{B} \) is countably generated, let \( \mathcal{B}_0 \) be a countable set of generators, and let \( \mathcal{B}_1 \) be the Boolean algebra generated by \( \mathcal{B}_0 \) under the usual set-theoretic Boolean operations. Then \( \mathcal{B}_1 \) is also countable. For each of the countably many \( D \in \mathcal{B}_1 \), let \( \mathcal{D}^D \) be a sequence of countable measurable partitions causing \( P^+_n(-, D) \) to converge to \( P(-, D) \) as described above. Let \( D_n \) be a sequence that \( \subseteq \)-majorizes all the \( \mathcal{D}^D_n \); for definiteness, we can take \( D_n = \bigsqcup_{m=1}^{\infty} D^B_m \), where \( B_0, B_1, \ldots \) is an enumeration of \( \mathcal{B}_1 \). Let \( F = \bigcap_n F_{D_n} \).

**Lemma 3.4** \( P^+_n(s, -) \) is countably additive on \( \mathcal{B}_1 \).

**Proof.** Let \( \{D_m \mid m \geq 0\} \) be a countable family of pairwise disjoint elements of \( \mathcal{B}_1 \) whose union is in \( \mathcal{B}_1 \). Since countable sums with suprema for bounded sets of nonnegative reals, we have
\[
P^+_n(s, \bigcup_m D_m) = \sum_{B \in \mathcal{D}_m} \sup_{\mu(B) > 0, \mu(C) > 0} \frac{\theta(C \times D_m)}{\mu(C)} \cdot 1_X(s, B) = \sum_{B \in \mathcal{D}_m} \sup_{\mu(B) > 0, \mu(C) > 0} \frac{\theta(C \times D_m)}{\mu(C)} \cdot 1_X(s, B) = \sum_{m} P^+_n(s, D_m).
\]

**Lemma 3.5** Let \( \{D_m \mid m \geq 0\} \) be a countable family of pairwise disjoint elements of \( \mathcal{B}_1 \) whose union is in \( \mathcal{B}_1 \). For all \( \varepsilon > 0 \), there exists \( N_\varepsilon \) such that for all \( n \),
\[
P^+_n(s, \bigcup_{m=N_\varepsilon}^{\infty} D_m) \leq \varepsilon.
\]

**Proof.** By Lemma 3.4, we have
\[
P^+_0(s, \bigcup_{m=1}^{\infty} D_m) = \sum_{m} P^+_0(s, D_m) = \sup_{N} \sum_{m=0}^{N-1} P^+_0(s, D_m),
\]

Similarly,
so for sufficiently large $N_\varepsilon$,

$$P_0^+(s, \bigcup_{m=N_\varepsilon}^\infty D_m) = \sum_{m=N_\varepsilon}^\infty P_0^+(s, D_m) \leq \varepsilon.$$ 

Moreover, for any $n \geq 0$,

$$P_n^+(s, \bigcup_{m=N_\varepsilon}^\infty D_m) \leq P_0^+(s, \bigcup_{m=N_\varepsilon}^\infty D_m) \leq \varepsilon.$$

By continuity of addition, the infimum operation on sets of nonnegative reals commutes with finite sum. This is not true for countable sums in general, but the following lemma shows that it is true in the special case at hand.

**Lemma 3.6** Let $\{D_m \mid m \geq 0\}$ be a countable family of pairwise disjoint elements of $B_1$ whose union is in $B_1$. For all $s \in F$,

$$\sum_m P(s, D_m) = \inf_n \sum_m P_n^+(s, D_m).$$

**Proof.** For the forward inequality, we have that for any $n$,

$$P(s, D_m) = \inf_n P_n^+(s, D_m) \leq P_n^+(s, D_m),$$

therefore by monotonicity of sum, $\sum_m P(s, D_m) \leq \sum_m P_n^+(s, D_m)$. As $n$ was arbitrary, $\sum_m P(s, D_m) \leq \inf_n \sum_m P_n^+(s, D_m)$.

For the reverse inequality, we have for any $\varepsilon > 0$ that

$$\sum_m P_n^+(s, D_m) = \sum_{m=0}^{N_\varepsilon-1} P_n^+(s, D_m) + \sum_{m=N_\varepsilon}^\infty P_n^+(s, D_m),$$

where $N_\varepsilon$ is from Lemma 3.5. By that lemma,

$$\inf_n \sum_m P_n^+(s, D_m) = \inf_{n=0}^{N_\varepsilon-1} \sum_m P_n^+(s, D_m) + \inf_{n=N_\varepsilon}^\infty \sum_m P_n^+(s, D_m)$$

$$= \sum_{m=0}^{N_\varepsilon-1} \inf_n P_n^+(s, D_m) + \inf_{n=N_\varepsilon}^\infty \sum_m P_n^+(s, \bigcup_{m=N_\varepsilon}^\infty D_m) \leq \sum_m P(s, D_m) + \varepsilon.$$ 

As $\varepsilon$ was arbitrary, the desired inequality holds. \qed

**Lemma 3.7** For all $s \in F$, $P(s, -)$ is countably additive on $B_1$.

**Proof.** Let $\{D_m \mid m \geq 0\}$ be a countable family of pairwise disjoint elements of $B_1$ whose union is in $B_1$. By Lemma 3.6,

$$P(s, \bigcup_m D_m) = \inf_n P_n^+(s, \bigcup_m D_m) = \inf_n \sum_m P_n^+(s, D_m) = \sum_m P(s, D_m).$$ 

\qed

**Proof of Theorem 3.1.** Lemma 3.7 establishes that the measurable functions $P(s, -)$ are defined and countably additive on $B_1$ for all $s \in F$, thus $P(s, -)$ is a pre-measure on $B_1$ for all $s \in F$. By the Carathéodory-Hahn-Kolmogorov extension theorem, $P(s, -)$ extends uniquely to a probability measure on $B$ for all $s \in F$. 

8
For \( s \notin F \), we can take \( P(s, -) \) to be any fixed measure on \( Y \), say \( P(s, B) = \theta(X \times B) \) or \( \theta(A \times B)/\theta(A \times Y) \) for any \( A \) such that \( \theta(A \times Y) \geq 0 \). The function \( P(-, B) \) is still a measurable function for any fixed \( B \in \mathcal{B} \). This proves that \( P \) is a Markov kernel.

It remains to argue uniqueness. For all \( D \in \mathcal{B}_1 \), the function \( P(-, D) \), being a Radon-Nikodym derivative, is measurable and unique in the sense that if \( Q(-, D) \) also satisfies (7), then \( P(-, D) \) and \( Q(-, D) \) are defined and equal outside a \( \mu \)-nullset \( N_D \in \mathcal{A} \) (which can depend on \( Q \)). Moreover, this is true uniformly for all \( D \in \mathcal{B}_1 \) outside the \( \mu \)-nullset \( N = \bigcup_{D \in \mathcal{B}_1} N_D \in \mathcal{A} \). But if \( P(s, D) = Q(s, D) \) for all \( s \notin N \) and \( D \in \mathcal{B}_1 \), then the same is true for all \( D \in \mathcal{B} \), since \( \mathcal{B} \) is generated by \( \mathcal{B}_1 \). Thus \( P \) is uniquely determined modulo \( \equiv_{\mu} \).

Theorem 3.1 says that there is a function \( \mathbf{J}^{-1} \) that, given a joint distribution \( \theta \) on \( X \times Y \) with left marginal \( \mu \), produces a unique \( \equiv_{\mu} \)-equivalence class \( [P]_{\mu} \) of kernels such that (7) holds for any \( P \) chosen from the class, thus \( \mathbf{J}[P]_{\mu} = \theta \). It follows that \( \mathbf{J} \) and \( \mathbf{J}^{-1} \) are inverses.

Given a kernel \( P : X \to Y \) and a probability measure \( \mu \), let \( \nu = \mu ; P \). Theorem 3.1 guarantees the existence of a kernel \( P^\dagger : Y \to X \) such that

\[
\mathbf{J} P^\dagger (B \times A) = \mathbf{J} P(A \times B),
\]

and \( P^\dagger \) is unique modulo \( \equiv_{\nu} \). We also have \( \mu = \nu ; P^\dagger \), thus \( \mu = \mu ; P ; P^\dagger, \nu = \nu ; P^\dagger ; P \), and \( P^\dagger \equiv_{\nu} P \).

It is not necessarily the case that \( P ; P^\dagger \equiv_{\nu} 1_X \), nor \( P^\dagger ; P \equiv_{\nu} 1_Y \); but the latter holds if \( P \) is deterministic.

This is the essential content of the standard disintegration theorem, which we treat next.

## 4 Standard Disintegration

Let \((X, A)\) and \((Y, B)\) be countably generated measurable spaces. Every measurable function \( f : X \to Y \) gives rise to a deterministic kernel \( P_f : X \to Y \) defined by

\[
P_f(s, B) = 1_Y(f(s), B) = 1_X(s, f^{-1}(B)).
\]

Modulo \( \equiv_{\mu} \), this is a morphism \( P_f : (X, A, \mu) \to (Y, B, \nu) \) of \( \text{Krn} \), provided \( \nu \) is the push-forward measure \( \nu = \mu \circ f^{-1} \); equivalently, if \( \nu = \mu ; P_f \). The joint distribution associated with \( P_f \) is

\[
\mathbf{J} P_f (A \times B) = \int_A 1_X(s, f^{-1}(B)) \mu(ds) = \mu(A \cap f^{-1}(B))
\]

with marginals \( \mu \) and \( \nu \).

The standard disintegration theorem, as presented for example in [2, 5, 9], essentially says that there exists a kernel \( Q : Y \to X \) that is a right inverse \(^5\) of \( P_f \) modulo \( \equiv_{\nu} \), in the sense that \( Q : P_f \equiv_{\nu} 1_Y \). Moreover, \( Q \) is unique up to \( \equiv_{\nu} \). Typically stronger assumptions than those required for Theorem 3.1 are assumed. In this section we show that the traditional version is a corollary of Theorem 3.1.

**Corollary 4.1** Given a measurable function \( f : X \to Y \) on countably generated spaces and a probability measure \( \mu \) on \( X \), let \( \nu = \mu \circ f^{-1} \) be the push-forward measure on \( Y \). The kernel \( P_f^\dagger : (Y, B, \nu) \to (X, A, \mu) \) is a right inverse of the deterministic kernel \( P_f \) modulo \( \equiv_{\nu} \); that is, \( P_f^\dagger ; P_f \equiv_{\nu} 1_Y \).

**Proof.** Observing that

\[
(P_f^\dagger ; P_f)(t, C) = \int_X P_f^\dagger(t, ds) \cdot 1_X(s, f^{-1}(C)) = P_f^\dagger(t, f^{-1}(C))
\]

and using (17) and (18), we have for all \( B, C \in \mathcal{B} \) that

\[
(\nu ; B ; P_f^\dagger ; P_f)(C) \equiv (\nu ; B ; P_f^\dagger)(f^{-1}(C)) = \mathbf{J} P_f^\dagger (B \times f^{-1}(C)) \equiv \mathbf{J} P_f (f^{-1}(C) \times B) = \mu(f^{-1}(C) \cap f^{-1}(B)) = \mu(f^{-1}(B \cap C)) = \nu(B \cap C) = (\nu ; B)(C),
\]

therefore \( P_f^\dagger ; P_f \equiv_{\nu} 1_Y \). \( \square \)

\(^5\) Since we write compositions in diagrammatic order \( Q ; P \), right inverses actually appear on the left.
5 Krn and JDist

To show that JDist is a category and that $J : \text{Krn} \to \text{JDist}$ is a functor, we must define composition and the identity morphisms in JDist and show that they are preserved by J.

Composition in JDist is defined as follows. For $\theta : (X, A, \mu) \to (Y, B, \nu)$ and $\eta : (Y, B, \nu) \to (Z, C, \xi)$,

$$((\theta ; \eta) (A \times C) = \lim_{\mathcal{D}} \sum_{B \in \mathcal{D}, \nu(B) > 0} \frac{\theta(A \times B) \cdot \eta(B \times C)}{\nu(B)}, \tag{19}$$

where the limit is taken over countable measurable partitions $\mathcal{D}$ of the mediating space $Y$. We argue below (Theorem 5.1) that the limit exists. The identity morphisms are the joint distributions $J1_X$ obtained from the identity kernels $1_X : X \to X$.

Note that this definition is completely symmetric in the input and output space. The category JDist is thus a dagger category whose involution $^\dagger$ is composition with transpose: $\theta^\dagger(C \times A) = \theta(A \times C)$.

**Theorem 5.1** Let $\mu, \nu, \xi$ be finite measures on $Y$. Let $(d\nu/d\mu)_D^+$ be the approximants defined for $\nu$ and $\mu$ as in §3.1, and let $(d\xi/d\mu)^+_D$ be those defined for $\xi$ and $\mu$. If there exists a countable measurable partition $D$ such that

$$\int_Y \left( \frac{d\nu}{d\mu} \right)_D^+ (t) \left( \frac{d\xi}{d\mu} \right)_D^+ (t) \mu(dt) < \infty,$$

then the limit

$$\lim_{\mathcal{D}} \sum_{B \in \mathcal{D}, \mu(B) > 0} \frac{\nu(B) \cdot \xi(B)}{\mu(B)} \tag{20}$$

exists and is equal to

$$\int_Y f(t) g(t) \mu(dt) = \inf_n \int_Y \left( \frac{d\nu}{d\mu} \right)_{D_n}^+ (t) \left( \frac{d\xi}{d\mu} \right)_{D_n}^+ (t) \mu(dt) \tag{21}$$

for any sufficiently fine countable $\sqsubseteq$-chain $D_0 \sqsubseteq D_1 \sqsubseteq \cdots$, where $f = \inf_n (d\nu/d\mu)^+_{D_n}$ and $g = \inf_n (d\xi/d\mu)^+_{D_n}$.

**Proof.** By definition of $(d\nu/d\mu)^+_D$ and $(d\xi/d\mu)^+_D$,

$$\int_Y \left( \frac{d\nu}{d\mu} \right)_D^+ (t) \left( \frac{d\xi}{d\mu} \right)_D^+ (t) \mu(dt) = \int_Y \left( \sum_{B \in \mathcal{D}, \mu(B) > 0} \frac{\nu(C)}{\mu(C)} 1_Y (t, B) \right) \left( \sum_{B \in \mathcal{D}, \mu(B) > 0} \frac{\xi(D)}{\mu(D)} 1_Y (t, B) \right) \mu(dt)$$

$$= \sum_{B \in \mathcal{D}, \mu(B) > 0} \frac{\nu(C)}{\mu(C)} \left( \sup_{D \subseteq B} \frac{\xi(D)}{\mu(D)} \right) \mu(B).$$

To show (20) and (21), for arbitrarily small positive $\varepsilon$, we have by Lemma 3.2

$$\sum_{B \in \mathcal{D}, \mu(B) > 0} \frac{\nu(B) \xi(B)}{\mu(B)} \leq \int_Y \left( \frac{d\nu}{d\mu} \right)^+_D (t) \left( \frac{d\xi}{d\mu} \right)^+_D (t) \mu(dt)$$

$$\leq \sum_{B \in \mathcal{D}, \mu(B) > 0} \frac{\nu(C)}{\mu(C)} + \varepsilon \left( \sum_{D \subseteq B} \frac{\xi(D)}{\mu(D)} + \varepsilon \right) \mu(B) \leq \sum_{B \in \mathcal{D}, \mu(B) > 0} \left( \frac{\nu(B)}{\mu(B)} + \varepsilon \right) \left( \frac{\xi(B)}{\mu(B)} + \varepsilon \right) \mu(B)$$

$$\leq \sum_{B \in \mathcal{D}, \mu(B) > 0} \left( \frac{\nu(B) \xi(B)}{\mu(B)} + \varepsilon \xi(B) + \varepsilon \nu(B) + \varepsilon^2 \mu(B) \right) \leq \left( \sum_{B \in \mathcal{D}, \mu(B) > 0} \frac{\nu(B) \xi(B)}{\mu(B)} \right) + \varepsilon \xi(Y) + \varepsilon \nu(Y) + \varepsilon^2 \mu(Y).$$
For the remaining statement (21), we use the stronger claim (10) of Lemma 3.2. Since

\[
\left( \frac{dw}{d\mu} \right)_D (t) \leq f(t) \leq \left( \frac{dw}{d\mu} \right)_D^+ (t), \quad \left( \frac{d\xi}{d\mu} \right)_D (t) \leq g(t) \leq \left( \frac{d\xi}{d\mu} \right)_D^+ (t),
\]

we have

\[
\left( \frac{dw}{d\mu} \right)_D (t) \left( \frac{d\xi}{d\mu} \right)_D (t) \leq \frac{d}{dt} \left( \frac{d\xi}{d\mu} \right)_D (t).
\]

We must choose \( D \) so that

\[
\left( \frac{dw}{d\mu} \right)_n (t) \left( \frac{d\xi}{d\mu} \right)_n^+ (t) - \left( \frac{dw}{d\mu} \right)_n (t) \left( \frac{d\xi}{d\mu} \right)_n^- (t)
\]

becomes arbitrarily small. By (10) of Lemma 3.2, \( D \) can be chosen so that

\[
\left( \frac{dw}{d\mu} \right)_D (t) \left( \frac{d\xi}{d\mu} \right)_D (t) - \left( \frac{dw}{d\mu} \right)_D (t) \left( \frac{d\xi}{d\mu} \right)_D (t) \leq 2\varepsilon^2.
\]

We should also argue that \( \theta ; \eta \) as defined in (19) on measurable rectangles extends to a joint probability measure on \( X \times Z \). This can be done using the Carathéodory-Hahn-Kolmogorov extension theorem. It suffices to verify the premises of that theorem, namely

(i) \( \theta ; \eta \) is finitely additive on measurable rectangles: if \( \{ A_n \times C_n \}_n \) is a finite set of pairwise disjoint measurable rectangles whose union is a measurable rectangle, then

\[
(\theta ; \eta)(\bigcup_n (A_n \times C_n)) = \sum_n (\theta ; \eta)(A_n \times C_n).
\]

(ii) If for each \( i \geq 0 \), \( \{ A_{n+i} \times C_{n+i} \}_n \) is a finite collection of pairwise disjoint measurable rectangles with

\[
\bigcup_n (A_{n+i} \times C_{n+i}) \subseteq \bigcup_n (A_{n} \times C_{n}),
\]

and if \( \bigcap_i \bigcup_n (A_{n+i} \times C_{n+i}) = \emptyset \), then

\[
\inf_i (\theta ; \eta)(\bigcup_n (A_{n+i} \times C_{n+i})) = 0.
\]

For (i), we can assume without loss of generality that the \( A_n \) are pairwise disjoint and the \( C_m \) are pairwise disjoint, and we are to show

\[
(\theta ; \eta)(\bigcup_n A_n \times \bigcup_m C_m) = \sum_n \sum_m (\theta ; \eta)(A_n \times C_m).
\]

Since limits commute with finite sums,

\[
(\theta ; \eta)(\bigcup_n A_n \times \bigcup_m C_m) = \lim_D \sum_{B \in D} \frac{\theta(\bigcup_n A_n \times B) \cdot \eta(B \times \bigcup_m C_m)}{\nu(B)} = \lim_D \sum_{B \in D} \sum_n \sum_m \frac{\theta(A_n \times B) \cdot \eta(B \times C_m)}{\nu(B)}
\]

\[
= \sum_n \sum_m \lim_D \sum_{B \in D} \frac{\theta(A_n \times B) \cdot \eta(B \times C_m)}{\nu(B)} = \sum_n \sum_m (\theta ; \eta)(A_n \times C_m).
\]

For (ii),

\[
\inf_i (\theta ; \eta)(\bigcup_n (A_{n+i} \times C_{n+i})) = \inf_i \sum_n (\theta ; \eta)(A_{n+i} \times C_{n+i})
\]

\[
= \inf_i \sum_n \lim_D \sum_{B \in D} \frac{\theta(A_{n+i} \times B) \cdot \eta(B \times C_{n+i})}{\nu(B)} = \inf_i \lim_D \sum_{B \in D} \sum_n \frac{\theta(A_{n+i} \times B) \cdot \eta(B \times C_{n+i})}{\nu(B)}.
\]
We argue that if \( \bigcap_i \bigcup_n (A^i_n \times C^i_n) = \emptyset \), then either \( \bigcap_i \bigcup_n A^i_n = \emptyset \) or \( \bigcap_i \bigcup_n C^i_n = \emptyset \). Suppose not. Let \( s \in \bigcap_i \bigcup_n A^i_n \) and \( t \in \bigcap_i \bigcup_n C^i_n \). Then for all \( i \) there exists \( n \) such that \( s \in A^i_n \) and there exists \( m \) such that \( t \in C^i_m \). By renumbering if necessary, we can assume that for all \( i \), \( s \in A^i_1 \) and \( t \in C^i_1 \). Then \( (s, t) \in \bigcap_i (A^i_1 \times C^i_1) \subseteq \bigcap_i \bigcup_n (A^i_n \times C^i_n) \).

By symmetry, assume without loss of generality that \( \bigcap_i \bigcup_n A^i_n = \emptyset \). Since \( \eta(B \times C^i_n)/\nu(B) = \eta(B \times C^i_n)/\eta(B \times Z) \leq 1 \), we have

\[
\inf_i \lim_D \sum_{B \in D} \sum_n \frac{\theta(A^i_n \times B) \cdot \eta(B \times C^i_n)}{\nu(B)} \leq \inf_i \lim_D \sum_n \theta(A^i_n \times B) = \inf \left( \bigcup_i A^i_n \times Y \right) = 0.
\]

**Theorem 5.2** The map \( J \) constitutes a full and faithful embedding \( J : \text{Krn} \to \text{JDist} \). Thus \( \text{Krn} \) is isomorphic to the full subcategory of \( \text{JDist} \) on countably generated spaces.

**Proof.** Composition in \( \text{Krn} \) is defined by \([P]_\mu \cdot [Q]_\nu = [P \cdot Q]_{\mu \cdot \nu}\). This is well defined by Lemma 2.1. We must show that \( J(P \cdot Q) = JP \cdot JQ \) and that the \( J \) and the \( J \) are identities for composition in \( \text{JDist} \).

For composition, using the fact that for any \( P, JP(A \times B) = (\mu ; A ; P)(B) \), the left-hand side gives

\[
J(P \cdot Q)(A \times C) = (\mu ; A ; P)Q(C).
\]

For the right-hand side, we observe that

\[
\inf_{t \in B} Q(t, C) \int_B \nu(dt) \leq \int_B Q(t, C) \nu(dt) \leq \sup_{t \in B} Q(t, C) \int_B \nu(dt),
\]

or in other words,

\[
\inf_{t \in B} Q(t, C) \nu(B) \leq (\nu ; B ; Q)(C) \leq \sup_{t \in B} Q(t, C) \nu(B),
\]

so for \( \nu(B) > 0 \),

\[
\inf_{t \in B} Q(t, C) \leq \frac{(\nu ; B ; Q)(C)}{\nu(B)} \leq \sup_{t \in B} Q(t, C).
\]

Thus

\[
(JP \cdot JQ)(A \times C) = \lim_D \sum_{B \in D} \frac{JP(A \times B) \cdot JQ(B \times C)}{\nu(B)} = \lim_D \sum_{B \in D, \mu(B) > 0} \frac{(\mu ; A ; P)(B) \cdot (\nu ; B ; Q)(C)}{\nu(B)}
\]

\[
= \lim_D \sum_{B \in D, \mu(B) > 0} (\mu ; A ; P)(B) \cdot \sup_{t \in B} Q(t, C) = \int_Y (\mu ; A ; P)(dt) \cdot Q(t, C) = (\mu ; A ; P ; Q)(C).
\]

For the identities, \( J1_X ; JP = J(1_X ; P) = JP \) and \( JP ; J1 = J(P ; 1_Y) = JP \). \( \square \)

6 Conclusion

We have given a version of disintegration based on weaker assumptions than previously known. The main technical contribution is an enhanced version of the Radon-Nikodym theorem that makes explicit the approximating functions, thereby allowing the coordinated construction of multiple simultaneous derivatives. We have used this to define a category \( \text{JDist} \) of joint distributions in which a category \( \text{Krn} \) of Markov kernels on countably generated spaces is fully and faithfully embedded.

For the future, we would like to use these results to formulate a point-free treatment based on the observation that the objects \( (X, A, \mu) \) of \( \text{Krn} \) and \( \text{JDist} \) do not really depend on the measure \( \mu \), but only its \( \sigma \)-ideal of nullsets.
References


