HILBERT SPACES AND THE RADON-NIKODYM THEOREM

STEVEN P. LALLEY

1. DEFINITIONS

Definition 1. A real inner product space is a real vector space V together with a *symmetric, bilinear, positive-definite* mapping $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$, that is, a mapping such that

$$\langle x, y\rangle = \langle y, x\rangle, \tag{1}$$

$$\langle ax + bx', y\rangle = a\langle x, y\rangle + b\langle x', y\rangle, \tag{2}$$

$$\langle x, x\rangle > 0 \quad \text{for all } x \neq 0. \tag{3}$$

A complex inner product space is a complex vector space V together with a *Hermitian, positive-definite* mapping $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$, that is, a mapping such that

$$\langle x, y\rangle = \overline{\langle y, x\rangle}, \tag{4}$$

$$\langle ax + bx', y\rangle = a\langle x, y\rangle + b\langle x', y\rangle, \tag{5}$$

$$\langle x, x\rangle > 0 \quad \text{for all } x \neq 0, \tag{6}$$

where the bar in the first equation denotes complex conjugation. In either case, for any $x \in V$ define

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Proposition 1. *(Cauchy-Schwarz Inequality)* $|\langle x, y\rangle| \leq \|x\|\|y\|$.

Proof. The proof relies on a common trick in Hilbert space theory called *polarization*, which is used to deduce information about inner products $\langle x, y\rangle$ from the positivity of inner products $\langle ax + by, ax + by\rangle$. It suffices to consider the special case where $\|x\| = \|y\| = 1$ (why?). For any scalars $a, b \in \mathbb{C}$,

$$\begin{aligned} 0 \leq \|ax + by, ax + by\| &= |a|^2\|x\|^2 + |b|^2\|y^2\| + 2\Re(ab\langle x, y\rangle) \\ &= |a|^2 + |b|^2 + 2\Re(ab\langle x, y\rangle) \end{aligned}$$

Set $a = 1$ and choose $b \in \mathbb{C}$ such that $|b| = 1$ and $\Re(ab\langle x, y\rangle) = -|\langle x, y\rangle|$; then

$$\begin{aligned} -2\Re(ab\langle x, y\rangle) \leq |a|^2 + |b|^2 \quad &\Longrightarrow \\ 2|\langle x, y\rangle| \leq 1 + 1 \quad &\Longrightarrow \\ |\langle x, y\rangle| \leq \|x\|\|y\|. \end{aligned}$$

□

Proposition 2. $\|\cdot\|$ *is a* norm *on* V*, that is, it satisfies the triangle inequality* $\|x + y\| \leq \|x\| + \|y\|$.

Proof. Exercise. □

Definition 2. A real (respectively, complex) *Hilbert space* is a real (respectively, complex) inner product space that is a *complete* metric space relative to the metric $d(x,y) := \|y-x\|$.

Example 1. If V is a linear subspace of a Hilbert space then its closure $\bar{V}$ (with respect to the metric d) is a Hilbert space.

Example 2. For any measure space $(\Omega,\mathscr{F},\mu)$ the space $L^2(\Omega,\mathscr{F},\mu)$ consisting of all square-integrable, measurable functions f on Ω (where f and g are identified if $f=g$ a.e.) is a Hilbert space when endowed with the inner product

$$\langle f,g\rangle = \int fg\,d\mu \quad (\text{real } L^2),$$
$$\langle f,g\rangle = \int f\overline{g}\,d\mu \quad (\text{complex } L^2).$$

NOTE: (1) That $L^2(\Omega,\mathscr{F},\mu)$ is *complete* follows from the Chebyshev inequality and the Borel-Cantelli Lemma. See 381 notes. (2) In what follows it will be important to remember that the $\sigma-$algebra $\mathscr{F}$ plays an important role in the definition: in particular, if $\mathscr{G}$ is a $\sigma-$algebra contained in $\mathscr{F}$ then $L^2(\Omega,\mathscr{G},\mu)$ is a closed linear subspace of $L^2(\Omega,\mathscr{F},\mu)$.

Example 3. If $\mathscr{Y}$ is a countable set then the vector space of all square-summable functions $f:\mathscr{Y}\to\mathbb{F}$, where either $\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$, is a Hilbert space, sometimes denoted $\ell^2(\mathscr{Y})$. Note: This is a special case of Example 2, with μ the counting measure on $\mathscr{Y}$.

2. ORTHONORMAL SETS AND ORTHONORMAL BASES

Definition 3. Two vectors $x,y\in H$ are *orthogonal* if their inner product is 0. A vector x is a *unit vector* if its norm is 1. A set of vectors $\{x_\lambda\}_{\lambda\in\Lambda}$ is an *orthonormal set* if each x_λ is a unit vector, and if any two distinct x_λ, x_ν are orthogonal. An *orthonormal basis* of H is an orthonormal set $\{x_\lambda\}_{\lambda\in\Lambda}$ such that the set of all *finite* linear combinations $\sum_{i=1}^m a_i e_{\lambda_i}$ is dense in H.

Fact 1. *If $\{e_k\}_{k\le n}$ is a finite orthonormal set then for any choice of scalars a_k and b_k,*

$$\left\langle \sum_{k=1}^n a_k e_k, \sum_{k=1}^n b_k e_k \right\rangle = \sum_{k=1}^n a_k\overline{b_k} \quad \textit{and} \tag{7}$$

$$\left\| \sum_{k=1}^n a_k e_k \right\|^2 = \sum_{k=1}^n |a_k|^2. \tag{8}$$

Definition 4. Let $\{e_k\}_{k\le n}$ be a finite orthonormal set and denote by V their linear span. Define the *orthogonal projection* operator $P_V : H\to V$ by

$$P_V x = \sum_{k=1}^n \langle x,e_k\rangle e_k \tag{9}$$

Fact 2. *If $\{e_k\}_{k\le n}$ is a finite orthonormal set with ($n-$dimensional) linear span V then for any $x\in H$,*

$$\|x-P_Vx\| = \min_{y\in V}\|x-y\|; \tag{10}$$

$$\langle (x-P_Vx), y\rangle = 0 \quad \textit{for all } y\in V; \tag{11}$$

$$\|x\|^2 = \|P_Vx\|^2 + \|x-P_Vx\|^2. \tag{12}$$

Since the linear span of x and the vectors e_k is a finite-dimensional inner product space, these facts all follow from high-school linear algebra. The third identity is a form of the *Pythagorean law*; a noteworthy consequence is the following inequality.

Corollary 1. *(Bessel Inequality) If $\{e_\lambda\}_{\lambda\in\Lambda}$ is an orthonormal set in a Hilbert space H, then for any $x \in H$*

$$\|x\|^2 \geq \sum_{\lambda\in\Lambda} |\langle x, e_\lambda\rangle|^2. \tag{13}$$

Consequently, at most countably many of the inner products $\langle x, e_\lambda\rangle$ are nonzero.

Proof. The sum on the right is by definition the supremum of the sums over finite subset of Λ, so the result follows from the Pythagorean law and equality (8). □

The identity (11) is the basis for the *Gram-Schmidt algorithm* for producing orthonormal bases. It implies that if x is not contained in V then an orthonormal basis for the linear span of $\{x\} \cup \{e_k\}_{k\leq n}$ can be obtained by adjoining the vector $(x - P_V x)/\|x - P_V x\|$ to the collection $\{e_k\}_{k\leq n}$. Induction then leads to the following fact.

Proposition 3. *(Gram-Schmidt) If $\{v_n\}_{n\in\mathbb{N}}$ is a countable set of linearly independent vectors in an inner product space then there exists a countable orthonormal set $\{u_n\}_{n\in\mathbb{N}}$ such that for every $m \in \mathbb{N}$ the span of $\{u_n\}_{n\leq m}$ is the same as the span of $\{v_n\}_{n\leq m}$.*

Proposition 4. *Every Hilbert space has an orthonormal basis. Moreover, if the Hilbert space is separable and infinite-dimensional then every orthonormal basis is countable.*

Proof for separable Hilbert spaces. If H is separable but infinite-dimensional then there exists a countable set $\{v_n\}_{n\in\mathbb{N}}$ of linearly independent vectors such that finite linear combinations of the vectors v_n are dense in H. (Start with a countable dense set, then inductively throw away vectors that are contained in the spans of the earlier vectors in the list.) Now apply Gram-Schmidt to the set $\{v_n\}_{n\in\mathbb{N}}$; the resulting orthonormal set $\{u_n\}_{n\in\mathbb{N}}$ is an orthonormal basis. That every orthonormal basis is countable is left as an exercise. □

Proof for arbitrary Hilbert spaces. Unfortunately, the proof for the general case seems to require the *Axiom of Choice*. Obviously, the set of orthonormal sets is partially ordered by inclusion. The *Hausdorff maximality principle*, one of the equivalent forms of the Axiom of Choice, asserts that every nonempty partially ordered set contains a *maximal chain*, that is, a maximal totally ordered subset. Let $\{e_\lambda\}_{\lambda\in\Lambda}$ be the union of all the sets in a maximal chain of orthonormal sets. This must be an orthonormal basis, because if not there would be a vector y not in the closure of the span of the e_λ, in which case the chain could be augmented by adding to the set $\{e_\lambda\}_{\lambda\in\Lambda}$ the vector

$$y - \sum_{\lambda\in\Lambda} \langle y, e_\lambda\rangle e_\lambda.$$

□

REMARK. Just about every *interesting* Hilbert space that you will ever encounter is separable, and therefore has either a finite or a countable orthonormal basis.

Proposition 5. *(Parseval's identities; orthogonal expansions) Let* $\{e_\lambda\}_{\lambda\in\Lambda}$ *be an orthonormal basis for the Hilbert space* H*. Then for any* $x \in H$*,*

$$x = \sum_{\lambda\in\Lambda} \langle x, e_\lambda\rangle e_\lambda \tag{14}$$

in the sense that if $\{e_k\}$ *is the countable (or finite) subset such that* $\langle x, e_\lambda\rangle \neq 0$ *then*

$$\lim_{n\to\infty} \|x - \sum_{k=1}^{n} \langle x, e_k\rangle e_k\| = 0.$$

Furthermore,

$$\langle x, y\rangle = \sum_{\lambda\in\Lambda} \langle x, e_\lambda\rangle \overline{\langle y, e_\lambda\rangle} \quad \forall\, x, y \in H \quad \textit{and} \tag{15}$$

$$\|x\|^2 = \sum_{\lambda\in\Lambda} |\langle x, e_\lambda\rangle|^2 \quad \forall x \in H. \tag{16}$$

Proof. Since $\{e_\lambda\}_{\lambda\in\Lambda}$ is an orthonormal basis, the set of *finite* linear combinations of vectors e_λ is dense in H. Consequently, for each $x \in H$ and any positive integer k there is a finite linear combination of the vectors e_λ is within distance 2^{-k} of x. Thus, there is a countable (or finite) subset $\{e_n\}_{n\in\mathbb{N}}$ such that x is in the closure of the linear span of the set $\{e_n\}_{n\in\mathbb{N}}$. By Fact 2 above, for each finite n the linear combination of $\{e_j\}_{j\leq n}$ closest to x is

$$P_V x = \sum_{j=1}^{n} \langle x, e_j\rangle e_j.$$

It now follows that

$$x = \lim_{n\to\infty} \sum_{j=1}^{n} \langle x, e_j\rangle e_j.$$

This and the Pythagorean identity (12) imply equation (16), and (12) also implies that there is no e_λ not contained in the countable set $\{e_n\}_{n\in\mathbb{N}}$ such that $\langle x, e_\lambda\rangle \neq 0$. The inner product formula (15) follows from (14), Fact 1, and the Cauchy-Schwarz inequality.

□

3. Consequences: Riesz-Fisher Theorem, Orthogonal Projections

3.1. **Bounded Linear Functionals.** Let H be a real (respectively, complex) Hilbert space. A *bounded linear functional* on H is a linear transformation $L : H \to \mathbb{R}$ (respectively, $\mathbb{C}$) such that for some $C < \infty$,

$$|Lx| \leq C\|x\| \quad \text{for every } x \in H.$$

The smallest C for which this inequality holds is called the *norm* of L, denoted $\|L\|$.

For any fixed vector $y \in H$, the mapping L defined by $Lx = \langle x, y\rangle$ is a bounded linear transformation, by the Cauchy-Schwarz inequality, and the norm is $\|L\| = \|y\|$. The *Riesz-Fisher theorem* states that there are no other bounded linear functionals. This fact is quite useful, as it can often be used to deduce the *existence* of vectors with desirable properties. See the proof of the Radon-Nikodym theorem later in the notes for an example of this principle.

Proposition 6. *(Riesz-Fisher) If L is a bounded linear functional on H then there is a unique $y \in H$ such that for every $x \in H$,*

$$Lx = \langle x, y \rangle.$$

Proof. Uniqueness is easily checked. Let $\{e_\lambda\}_{\lambda\in\Lambda}$ be an orthonormal basis, and define scalars a_λ by $\overline{a_\lambda} = Le_\lambda$. It must be the case that $\sum_\lambda |a_\lambda|^2 \leq \|L\| < \infty$, because for any finite subset $\{e_k\}_{k\leq n}$,

$$\left|L(\sum a_k e_k)\right| = \sum |a_k|^2 \leq \|L\| \|\sum a_k e_k\|.$$

Consequently, $y = \sum_\lambda a_\lambda e_\lambda$ is a well-defined element of H with norm $\leq C$. Parseval's identity (15) now implies that for each e_λ,

$$Le_\lambda = \langle e_\lambda, y \rangle.$$

This identity extends by linearity to finite linear combinations of the e_λ, and then by boundedness to all $x \in H$. □

3.2. **Orthogonal Projections.** Let V be a linear subspace of the Hilbert space H, and define $\overline{V}$ to be its closure (with respect to the norm on H). This is itself a linear subspace (why?) in which V is dense. Define $V^\perp$ to be the set of all vectors $x \in H$ that are *orthogonal* to V, i.e., such that $\langle x, y \rangle = 0$ for every $y \in V$. Observe that $V^\perp$ is closed, and

$$(V^\perp)^\perp = \overline{V} \quad \text{and}$$
$$V^\perp = (\overline{V})^\perp.$$

Proposition 7. *Let V be a closed linear subspace of a Hilbert space H and let $V^\perp$ be its orthogonal complement. If B_V is an orthonormal basis for V and $B_{V^\perp}$ is an orthonormal basis for $V^\perp$ then the union $U = B_V \cup B_{V^\perp}$ is an orthonormal basis for H. Consequently, every $y \in H$ has a unique representation*

$$y = P_V y + P_{V^\perp} y \tag{17}$$

where $P_V y \in V$ and $P_{V^\perp} y \in V^\perp$.

Proof. It is clear that U is an orthonormal *set*, because vectors in $V^\perp$ are by definition orthogonal to vectors in V. If U were not an orthonormal basis, then there would exist a vector $y \in H$ not in the closure of the linear span of U. Consequently,

$$z := y - \sum_\alpha \langle y, v_\alpha \rangle v_\alpha - \sum_\beta \langle y, w_\beta \rangle w_\beta \neq 0$$

But then z would be orthogonal to both V and $V^\perp$, which is a contradiction, because $V^\perp$ contains all vectors orthogonal to V.

Thus, U is an orthonormal basis for H. Consequently, for any $y \in H$,

$$y = \sum_{u\in B_V} \langle y, u \rangle u + \sum_{u \in B_{V^\perp}} \langle y, u \rangle u,$$

hence y is the sum of a vector in V and a vector in $V^\perp$. There can only be one way to do this, because if there were two different such sums both representing y then subtracting one from the other would yield an identity

$$0 = v + v^\perp \quad \text{with } v \in V \quad \text{and } v^\perp \in V^\perp,$$

which is impossible unless $v = v^\perp = 0$. □

The decomposition (17) shows that the Hilbert space H is, in effect, the *direct sum* of the Hilbert spaces V and $V^{\perp}$: in particular, every $x \in H$ can be associated with the pair (v, w) where $v = P_V x$ and $w = P_{V^{\perp}} x$, and inner products can be written as

$$\langle x, x' \rangle = \langle v, v' \rangle + \langle w, w' \rangle$$

where $x = v + w$ and $x' = v' + w'$.h The operator P_V in equation (17) is called the *orthogonal projection* onto V. It is the unique linear operator $P_V : H \to V$ satisfying

$$P_V x = x \quad \text{for all } x \in V \quad \text{and} \tag{18}$$

$$P_V x = 0 \quad \text{for all } x \in V^{\perp}. \tag{19}$$

Proposition 8. *(Nested Subspaces) Let H be a Hilbert space and let $V \subset W$ both be closed linear subspaces of H. Then*

$$P_V \circ P_W = P_V.$$

Proof. It suffices to show that $P_V \circ P_W$ satisfies the characteristic properties (18) – (19). The first is obvious. To prove the second, it suffices to show that if a vector $x \in H$ is orthogonal to V then $P_W x$ is also orthogonal to V.

Assume, then, that x is orthogonal to V, that is, $\langle x, y \rangle = 0$ for every $y \in V$. By Proposition 7, $x = P_W x + P_{W^{\perp}} x$. Clearly, if $y \in V$ then $y \in W$ and so $\langle P_{W^{\perp}} x, y \rangle = 0$. Consequently, by linearity,

$$\langle P_V x, y \rangle = \langle x, y \rangle - \langle P_{V^{\perp}} x, y \rangle = 0.$$

□

4. Radon-Nikodym Theorem

The Riesz-Fisher theorem can be used to give a relatively simple proof of one of the most useful theorems in measure theory, the *Radon-Nikodym* theorem.

Definition 5. Let μ, ν be two positive measures on a measurable space $(\Omega, \mathscr{F})$. The measure μ is said to be *absolutely continuous* relative to ν, written $\mu \ll \nu$, if every set of $\nu-$measure 0 is also a set of $\mu-$measure 0. The measures μ, ν are *mutually absolutely continuous* if each is absolutely continuous relative to the other, and are *mutually singular* if there exists a measurable set F such that $\mu(F) = 0$ and $\nu(F^c) = 0$.

Example 4. Let ν be a positive measure on $(\Omega, \mathscr{F})$ and let $h \in L^1(\Omega, \mathscr{F}, \nu)$ be a nonnegative, integrable function. For each $F \in \mathscr{F}$ define

$$\mu(F) := \int_F h \, d\nu.$$

Then μ is a measure, and $\mu \ll \nu$.

The Radon-Nikodym theorem states that if ν is a $\sigma-$finite measure, then every finite measure μ that is absolutely continuous relative to ν has the form of Example 4.

Theorem 1. *(Radon-Nikodym) Let ν be a $\sigma-$finite measure on $(\Omega, \mathscr{F})$ and let μ be a finite measure such that $\mu \ll \nu$. Then there exists an essentially unique, nonnegative function $h \in L^1(\Omega, \mathscr{F}, \nu)$ such that for every bounded, measurable function g on $(\Omega, \mathscr{F})$,*

$$\int g \, d\mu = \int g h \, d\nu. \tag{20}$$

The function h is called the Radon-Nikodym derivative *of μ with respect to ν, and sometimes denoted by*

$$h = \frac{d\mu}{d\nu}.$$

REMARK. When μ and ν are probability mesures, the term *likelihood ratio* is sometimes used in place of *Radon-Nikodym derivative*.

Proof. **Step 1: Uniqueness.** *Essential uniqueness* of the Radon-Nikodym derivative means up to sets of $\nu-$measure zero. Clearly, if one changes h on a set of measure 0 then none of the integrals on the right side of (20) is affected. To see that h must be essentially unique, observe that if h_1 and h_2 are two functions in $L^1(\Omega, \mathscr{F}, \nu)$ such that

$$\int \mathbf{1}_F h_1 \, d\nu = \int \mathbf{1}_F h_2 \, d\nu$$

for every set $F \in \mathscr{F}$, then the difference $h_1 - h_2$ integrates to 0 on every $F \in \mathscr{F}$, and therefore (why?) must equal 0 $\nu-$a.e.

Step 2: Reduction to Finite Measures. Next we show that it suffices to consider the special case where both μ and ν are probability measures. If both μ and ν are finite measures (i.e., assign finite mass to Ω) then both can be made into probability measures by multiplication by normalizing constants, and this change of normalization will not change the absolute continuity $\mu \ll \nu$. Now suppose that ν has infinite total mass but is $\sigma-$finite; then there exists a sequence of pairwise-disjoint set $F_n \in \mathscr{F}$ such that

$$\bigcup_{n=1}^{\infty} F_n = \Omega \quad \text{and} \quad 0 < \nu(F_n) < \infty \;\; \forall \, n \geq 1.$$

Define a probability measure λ as follows:

$$\lambda(B) = \sum_{n=1}^{\infty} 2^{-n} \frac{\nu(B \cap F_n)}{\nu(F_n)}.$$

It is not difficult to check that $\mu \ll \nu$ if and only if $\mu \ll \lambda$. Now suppose that the theorem is true for the pair μ, λ, that is, that for some nonnegative function $h \in L^1(\Omega, \mathscr{F}, \lambda)$,

$$\int g \, d\mu = \int g h \, d\lambda \quad \text{for all bounded } g.$$

Define

$$\tilde{h} = \sum_{n=1}^{\infty} (2^{-n}/\nu(F_n)) h \mathbf{1}_{F_n};$$

then it is easily checked that $\tilde{h} \in L^1(\Omega, \mathscr{F}, \nu)$ and that $\tilde{h}$ satisfies relation (20).

Step 3: Probability Measures. Finally, we use the Riesz-Fisher theorem to prove the theorem in the case where both μ and ν are probability measures. Define $\lambda = \mu + \nu$; this is a positive measure of total mass 2, and clearly both μ and ν are absolute continuous relative to λ. For $g \in L^2(\Omega, \mathscr{F}, \lambda)$, set

$$E_\mu(g) = \int g \, d\mu,$$
$$E_\nu(g) = \int g \, d\nu.$$

Then both E_μ and E_ν are bounded linear functionals on $L^2(\Omega, \mathscr{F}, \lambda)$. This (and the fact that E_μ and E_ν are well-defined on $L^2(\Omega, \mathscr{F}, \lambda)$) follows by the Cauchy-Schwartz (second moment) inequality: for E_μ,

$$\begin{aligned} |E_\mu(g)| &\le \left\{ \int g^2 \, d\mu \right\}^{1/2} \\ &\le \left\{ \int g^2 \, d\lambda \right\}^{1/2} \\ &= \|g\| \end{aligned}$$

where $\|g\|$ is the norm of g in the Hilbert space $L^2(\Omega, \mathscr{F}, \lambda)$. Consequently, Riesz-Fisher implies that there exist $h_1, h_2 \in L^2(\Omega, \mathscr{F}, \lambda)$ such that

$$E_\mu(g) = \int g h_1 \, d\lambda \quad \text{and}$$
$$E_\nu(g) = \int g h_2 \, d\lambda,$$

It is easy to see that both h_i are nonnegative, because the operators on the right side are expectations, and the two equations hold for all indicators $g = \mathbf{1}_F$. Moreover, both h_i must integrate to 1 against λ.

The last step will be to show that $h = h_1/h_2$ is the likelihood ratio of μ relative to ν. First note that the event $\{h_2 = 0\}$ is a set of $\nu-$measure 0, since by hypothesis $\mu \ll \nu$. Hence, $h_1/h_2 < \infty$ almost surely (with respect to ν). Now for any nonnegative, bounded g, by the monotone convergence theorem,

$$\begin{aligned} E_\mu g &= \int g h_1 \, d\lambda \\ &= \int g \frac{h_1}{h_2} h_2 \, d\lambda \\ &= \lim_{n\to\infty} \uparrow \int g \left(\frac{h_1}{h_2} \wedge n \right) h_2 \, d\lambda \\ &= \lim_{n\to\infty} \uparrow \int g \left(\frac{h_1}{h_2} \wedge n \right) d\nu \\ &= \int g \frac{h_1}{h_2} \, d\nu. \end{aligned}$$

NOTE: The reason for the truncation at n is to ensure that the integrand is in $L^2(\Omega, \mathscr{F}, \lambda)$, since we only know at this point that the likelihood ratio identity holds for square-integrable functions. □

At first glance it is easy to overlook the importance of the $\sigma-$algebra in the Radon-Nikodym theorem. A moment's thought will reveal, however, that even the absolute continuity relation depends critically on the choice of $\sigma-$algebra: two measures might be mutually absolutely continuous on one $\sigma-$algebra but singular one another. Following are two instructive examples.

Example 5. Let $\Omega = [0,1]$ be the unit interval, let $\mathscr{B}$ be the $\sigma-$algebra of Borel sets on Ω, and let $\mathscr{F}$ be the trivial $\sigma-$algebra $\{\emptyset, \Omega\}$. Let λ be Lebesgue measure, and let $\mu = \delta_0$ be the measure that puts all of its mass at the single point 0, so that for any $B \in \mathscr{B}$,

$$\begin{aligned} \mu(B) &= 1 \text{ if } 0 \in B, \\ &= 0 \text{ if } 0 \notin B. \end{aligned}$$

Then μ and λ are mutually singular on $\mathscr{B}$, but they are mutually absolutely continuous on $\mathscr{F}$.

Example 6. Let $\Omega = \{0,1\}^{\mathbb{N}}$ be the space of infinite sequences of 0s and 1s. For each $n \geq 1$ let $\mathscr{F}_n$ be the $\sigma-$algebra generated by the first n coordinate variables, and let $\mathscr{F}_\infty$ be the usual Borel field, that is, the smallest $\sigma-$algebra containing the union $\cup_{n\geq 1}\mathscr{F}_n$. For each $p \in (0,1)$ let P_p be the product Bernoulli-p measure on $(\Omega, \mathscr{F}_\infty)$, that is, the probability measure under which the coordinate random variables are i.i.d. Bernoulli-p. If $p \neq r$ then P_p and P_r are mutually singular on $\mathscr{F}_\infty$ (why?), but for any $n \in \mathbb{N}$ they are mutually absolutely continuous on $\mathscr{F}_n$ (what is the likelihood ratio?).

Example 6 illustrates another subtlety connected with the Radon-Nikodym theorem. The Radon-Nikodym theorem asserts that the Radon-Nikodym derivative is (essentially) unique, and that it is measurable relative to the $\sigma-$algebra in question. But the measures P_p, P_r of Example 6 are mutually absolutely continuous on each of the $\sigma-$algebras $\mathscr{F}_n$, so there is a likelihood ratio h_n for each n, measurable with respect to $\mathscr{F}_n$. Now if $m < n$ then $\mathscr{F}_m \subset \mathscr{F}_n$, and so it must be the case that for every $F \in \mathscr{F}_m$,

$$P_p(F) = E_r \mathbf{1}_F h_n.$$

Why doesn't this contradict the essential uniqueness of Radon-Nikodym derivatives? Because h_n isn't measurable with respect to $\mathscr{F}_m$! This is worth remembering: the Radon-Nikodym derivatives for a $\sigma-$algebra $\mathscr{F}$ is only unique within the class of $\mathscr{F}-$measurable functions.