The Radon-Nikodym theorem

(telegram style notes)

P.J.C. Spreij

this version: October 11, 2007

1 Linear functionals on $\mathbb{R}^n$

Let $E = \mathbb{R}^n$. It is well known that every linear map $T : E \to \mathbb{R}^m$ can uniquely be represented by an $m \times n$ matrix $M = M(T)$ via $Tx = Mx$, which we will prove below for the case $m = 1$. Take the result for granted, let $m = 1$ and $\langle \cdot, \cdot \rangle$ be the usual inner product on E, $\langle x, y \rangle = x^\top y$. For this case the matrix M becomes a row vector. Let $y = M^\top \in \mathbb{R}^n$, then we have

$$Tx = \langle x, y \rangle. \tag{1.1}$$

Hence we can identify the mapping T with the vector y. Let E^* be the set of all linear maps on E. Then we have for this case the identification of E^* with E itself via equation (1.1).

Suppose that we know that (1.1) holds. Then the kernel K of T is the space of vectors that are orthogonal to y and the orthogonal complement of K is the space of all vectors that are multiples of y. This last observation is the core of the following elementary proof of (1.1).

Let us first exclude the trivial situation in which $T = 0$. Let K be the kernel of T. Then K is a proper linear subspace of E. Take a nonzero vector z in the orthogonal complement of K. Every vector x can be written as a sum $x = \lambda x + u$, with $\lambda \in \mathbb{R}$ and $u \in K$. Then we have

$$Tx = \lambda Tz. \tag{1.2}$$

Of course we have

$$\lambda = \frac{\langle x, z \rangle}{\langle z, z \rangle}. \tag{1.3}$$

Let $y = \frac{Tz}{\langle z,z \rangle} z$. Then $\langle x, y \rangle = \frac{Tz}{\langle z,z \rangle} \langle x, z \rangle$. But then we obtain from (1.2) and (1.3) that $\langle x, y \rangle = Tx$. Uniqueness of y is shown as follows. Let $y' \in E$ be such that $Tx = \langle x, y' \rangle$. Then $\langle x, y - y' \rangle$ is zero for all $x \in E$, in particular for $x = y - y'$. But then $y - y'$ must be the zero vector.

The interesting observation is that this proof carries over to the case where one works with (continuous) linear functionals on a Hilbert space, which we treat in the next section.

2 Linear functionals on a Hilbert space

Let H be a (real) Hilbert space, a vector space over the real numbers, endowed with an inner product $\langle \cdot, \cdot \rangle$, that is complete w.r.t. the norm $|| \cdot ||$ generated by this inner product. Let T be a continuous linear functional on H. We will prove the *Riesz-Fréchet* theorem, which states that every continuous linear functional on H is given by an inner product with a fixed element of H.

Theorem 2.1 *There exists a unique element $y \in H$ such that $Tx = \langle x, y \rangle$.*

Proof. We exclude the trivial case in which $T = 0$. Let K be the kernel of T. Since T is linear, K is a closed subspace of H. Take an element w with $Tw \neq 0$. Since K is closed, the orthogonal projection u of w on K exists and we have $w = u + z$, where z belongs to the orthogonal complement of K. Obviously $z \neq 0$. The rest of the proof is exactly the same as in the previous section. □

This theorem can be summarized as follows. The dual space H^* of H (the linear space of all continuous linear functionals on H) can be identified with H itself. Moreover, we can turn H^* into a Hilbert space itself by defining an inner product $\langle \cdot, \cdot \rangle^*$ on H^*, Let $T, T' \in H^*$ and let y, y' the elements in H that are associate to H according to the theorem. Then we define $\langle T, T' \rangle^* = \langle y, y' \rangle$. One readily shows that this defines an inner product. Let $|| \cdot ||^*$ be the norm on H^*. Then H^* is complete as well. Indeed, let (T_n) be a Cauchy sequence in H^* with corresponding elements (y_n) in H, satisfying $T_n x \equiv \langle x, y_n \rangle$. Then $||T_n - T_m||^* = ||y_n - y_m||$. The sequence (y_n) is thus Cauchy in H and has a limit y. Define $Tx = \langle x, y \rangle$. Then T is obviously linear and $||T_n - T||^* = ||y_n - y|| \to 0$. Concluding, we say that the normed spaces $(H^*, || \cdot ||^*)$ and $(H, || \cdot ||)$ are isomorphic.

The usual *operator norm* of a linear functional T on a normed space is defined as $||T||^* = \sup_{x \neq 0} \frac{|Tx|}{||x||}$. It is a simple consequence of the Cauchy-Schwartz inequality that this norm $|| \cdot ||^*$ is the same as the one in the previous paragraph.

3 Real and complex measures

Consider a measurable space (S, Σ). A function $\mu : \Sigma \to \mathbb{C}$ is called a *complex* measure if it is countably additive. Such a μ is called a *real* or a *signed* measure if it has its values in $\mathbb{R}$. What we called a measure before, will now be called a *positive* measure. In these notes a measure is either a positive or a complex (or real) measure. Notice that a positive measure can assume the value infinity, unlike a complex measure, whose values lie in $\mathbb{C}$ (see also (3.4)).

Let μ be a complex measure and $E_1, E_2, \ldots$ be disjoint sets in Σ with $E = \bigcup_{i \geq 1} E_i$, then (by definition)

$$\mu(E) = \sum_{i \geq 1} \mu(E_i),$$

where the sum is convergent and the summation is independent of the order. Hence the series is absolutely convergent as well, and we also have

$$|\mu(E)| \leq \sum_{i \geq 1} |\mu(E_i)| < \infty. \tag{3.4}$$

For a given set $E \in \Sigma$ let $\Pi(E)$ be the collection of all *measurable* partitions of E, countable partitions of E with elements in Σ. If μ is a complex measure, then we define

$$|\mu|(E) = \sup\{\sum_i |\mu(E_i)| : E_i \in \pi(E) \text{ and } \pi(E) \in \Pi(E)\}.$$

It can be shown (and this is quite some work) that $|\mu|$ is a (positive) measure on (S, Σ) with $|\mu|(S) < \infty$ and it is called the *total variation measure* (of μ). Notice that always $|\mu|(E) \geq |\mu(E)|$ and that in particular $\mu(E) = 0$ as soon as $|\mu|(E) = 0$.

In the special case where μ is real valued,

$$\mu^+ = \frac{1}{2}(|\mu| + \mu)$$

and

$$\mu^- = \frac{1}{2}(|\mu| - \mu)$$

define two bounded positive measures such that

$$\mu = \mu^+ - \mu^-.$$

This decomposition of the real measure μ is called the Jordan decomposition.

4 Absolute continuity and singularity

Consider a measurable space (S, Σ). Let μ be a positive measure and λ a complex or positive measure on this space. We say that λ is *absolutely continuous* w.r.t. μ (notation $\lambda \ll \mu$), if $\lambda(E) = 0$ for every $E \in \Sigma$ with $\mu(E) = 0$. An example of absolute continuity we have seen already in the previous section: $\mu \ll |\mu|$ for a complex measure μ. The measures μ and λ are called *mutually singular* (notation $\lambda \perp \mu$) if there exist disjoint sets E and F in Σ such that $\lambda(A) = \lambda(A \cap E)$ and $\mu(A) = \mu(A \cap F)$ for all $A \in \Sigma$. Notice that in this case $\lambda(F) = \mu(E) = 0$.

Proposition 4.1 *Let μ be a positive measure and λ_1, λ_2 arbitrary measures, all defined on the same measurable space. Then the following properties hold true.*

1. *If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$.*
2. *If $\lambda_1 \ll \mu$ and $\lambda_2 \ll \mu$, then $\lambda_1 + \lambda_2 \ll \mu$.*
3. *If $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 \perp \lambda_2$.*
4. *If $\lambda_1 \ll \mu$ and $\lambda_1 \perp \mu$, then $\lambda_1 = 0$.*

Proof. Exercise 7.2. □

Proposition 4.2 *Let μ be a positive measure and λ_a and λ_s be arbitrary measures on (S, Σ). Assume that $\lambda_a \ll \mu$ and $\lambda_s \perp \mu$. Put*

$$\lambda = \lambda_a + \lambda_s. \tag{4.5}$$

Suppose that λ also admits the decomposition $\lambda = \lambda'_a + \lambda'_s$ with $\lambda'_a \ll \mu$ and $\lambda'_s \perp \mu$. Then $\lambda'_a = \lambda_a$ and $\lambda'_s = \lambda_s$.

Proof. It follows that

$$\lambda'_a - \lambda_a = \lambda_s - \lambda'_s,$$

$\lambda'_a - \lambda_a \ll \mu$ and $\lambda_s - \lambda'_s \perp \mu$ (proposition 4.1), and hence both are zero (proposition 4.1 again). □

The content of proposition 4.2 is that the decomposition (4.5) of λ, if it exists, is unique. We will see in section 5 that, given a positive measure μ, such a decomposition exists for any measure λ and it is called the *Lebesgue decomposition* of λ w.r.t. μ. Recall

Proposition 4.3 *Let μ be a positive measure on (S, Σ) and h a nonnegative measurable function on X. Then the map $\lambda : \Sigma \to [0, \infty]$ defined by*

$$\lambda(E) = \mu(1_E h) \tag{4.6}$$

is a positive measure on (S, Σ) that is absolutely continuous w.r.t. μ. If h is complex valued and in $\mathcal{L}^1(S, \Sigma, \mu)$, then λ is a complex measure.

Proof. See Williams, section 5.14 for nonnegative h. The other case is exercise 7.3. □

The Radon-Nikodym theorem of the next section states that every measure λ that is absolutely continuous w.r.t. μ is of the form (4.6). We will use in that case the notation

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

In the next section we use

Lemma 4.4 *Let μ be a finite positive measure and $f \in \mathcal{L}^1(S, \Sigma, \mu)$, possibly complex valued. Let A be the set of averages*

$$a_E = \frac{1}{\mu(E)} \int_E f \, d\mu,$$

where E runs through the collection of sets with $\mu(E) > 0$. Then $\mu(\{f \notin \bar{A}\}) = 0$.

Proof. Assume that $\mathbb{C} \setminus \bar{A}$ is not the empty set (otherwise there is nothing to prove) and let B be a closed ball in $\mathbb{C} \setminus \bar{A}$ with center c and radius $r > 0$. Notice that $|c - a| > r$ for all $a \in \bar{A}$. It is sufficient to prove that $E = f^{-1}[B]$ has measure zero, since $\mathbb{C} \setminus \bar{A}$ is a countable union of such balls.
Suppose that $\mu(E) > 0$. Then we would have

$$|a_E - c| \leq \frac{1}{\mu(E)} \int_E |f - c| \, d\mu \leq r.$$

But this is a contradiction since $a_E \in A$. □

5 The Radon-Nikodym theorem

The principal theorem on absolute continuity (and singularity) is

Theorem 5.1 *Let μ be a positive σ-finite measure and λ a complex measure. Then there exists a unique decomposition $\lambda = \lambda_a + \lambda_s$ and a function $h \in \mathcal{L}^1(S, \Sigma, \mu)$ (called the Radon-Nikodym derivative of λ_a w.r.t. μ and commonly denoted by $\frac{d\lambda_a}{d\mu}$) such that $\lambda_a(E) = \mu(1_E h)$ for all $E \in \Sigma$. Moreover, h is unique in the sense that any other h' with this property is such that $\mu(\{h \neq h'\}) = 0$.*

Proof. Uniqueness of the decomposition $\lambda = \lambda_a + \lambda_s$ is the content of proposition 4.2. Hence we proceed to show existence. Let us first assume that $\mu(S) < \infty$ and that λ is positive and finite.
Consider then the positive bounded measure $\phi = \lambda + \mu$. Let $f \in \mathcal{L}^2(S, \Sigma, \phi)$. The Schwartz inequality gives

$$|\lambda(f)| \leq \lambda(|f|) \leq \phi(|f|) \leq (\phi(f^2))^{1/2}(\phi(S))^{1/2}.$$

We see that the linear map $f \mapsto \lambda(f)$ is bounded on the pre-Hilbert space $\mathcal{L}^2(S, \Sigma, \phi)$. Hence there exists, by virtue of the Riesz-Fréchet theorem 2.1, a $g \in \mathcal{L}^2(S, \Sigma, \phi)$ such that for all f

$$\lambda(f) = \phi(fg). \tag{5.7}$$

Take $f = 1_E$ for any E with $\phi(E) > 0$. Then $\phi(E) \geq \lambda(E) = \phi(1_E g) \geq 0$ so that the average $\frac{1}{\phi(E)}\phi(1_E g)$ lies in $\in [0, 1]$. From lemma 4.4 we obtain that $\phi(\{g \notin [0, 1]\}) = 0$. Replacing g with $g1_{\{0 \leq g \leq 1\}}$, we see that (5.7) still holds and hence we may assume that $0 \leq g \leq 1$.
Take now $f = 1_B$, where $B = \{g = 1\}$. Then we obtain from (5.7) that $\lambda(\{g = 1\}) = \phi(\{g = 1\})$ and hence $\mu(\{g = 1\}) = 0$. Define then positive measures by $\lambda_a(E) = \lambda(E \cap B^c)$ and $\lambda_s(E) = \lambda(E \cap B)$. It is immediate that $\lambda = \lambda_a + \lambda_s$ and that $\lambda_s \perp \mu$.
Rewrite (5.7) as

$$\lambda((1 - g)f) = \mu(fg). \tag{5.8}$$

Let $A = B^c = \{g \in [0, 1)\}$, $E \in \Sigma$ and $n \geq 1$ be arbitrary and take $f = 1_{A \cap E}(1 + g + \cdots g^{n-1})$ in (5.8). Then we obtain

$$\lambda(1_{E \cap A}(1 - g^n)) = \mu(1_{E \cap A}(g + \cdots + g^n)).$$

The integral on the left converges by the dominated convergence theorem to $\lambda_a(E)$ and the integral on the right by the monotone convergence theorem to $\mu(1_E 1_A g/(1-g))$. Hence with the nonnegative function $h = 1_A g/(1-g)$ we have $\lambda_a(E) = \mu(1_E h)$, which is what he had to prove. Since $\mu(h) = \lambda_a(S) < \infty$, we also see that $h \in \mathcal{L}^1(S, \Sigma, \mu)$. Uniqueness of h is left as exercise 7.6.
If μ is not bounded but merely σ-additive and λ bounded and positive we decompose S into a measurable partition $S = \bigcup_{n \geq 1} S_n$, with $\mu(S_n) < \infty$. Apply the previous part of the proof to each of the spaces (S_n, Σ_n) with Σ_n

the trace σ-algebra of Σ on S_n. This yields measures $\lambda_{a,n}$ and functions h_n defined on the S_n. Put then $\lambda_a(E) = \sum_n \lambda_{a,n}(E \cap S_n)$, $h = \sum_n 1_{S_n} h_n$. Then $\lambda(E) = \mu(1_E h)$ and $\mu(h) = \lambda_a(S) < \infty$. For real measures λ we apply the results to λ^+ and λ^- and finally, if λ is complex we treat the real and imaginary part separately. The trivial details are omitted. □

Remark 5.2. If we take λ a positive σ-finite measure, then the Radon-Nikodym theorem is still true with the exception that we only have $\mu(h 1_{S_n}) < \infty$, where the S_n form a measurable partition of S such that $\lambda(S_n) < \infty$ for all n. Notice that in this case (inspect the proof above) we may take $h \geq 0$.

6 Additional results

Proposition 6.1 *Let μ be a complex measure. Then $\mu \ll |\mu|$ and the Radon-Nikodym derivative $h = \frac{d\mu}{d|\mu|}$ may be taken such that $|h| = 1$.*

Proof. Let h be any function as in the Radon-Nikodym theorem. Since $||\mu|(h 1_E)| = |\mu(E)| \leq |\mu|(E)$, it follows from lemma 4.4 that $|\mu|(\{|h| > 1\}) = 0$. On the other hand, for $A = \{|h| \leq r\}$ $(r > 0)$ and a measurable partition with elements A_j of A, we have

$$\sum_j |\mu(A_j)| = \sum_j |\mu|(1_{A_j} h) \leq \sum_j |\mu|(1_{A_j}|h|) \leq r|\mu|(A).$$

Then we find, by taking suprema over such partitions, that $|\mu|(A) \leq r|\mu|(A)$. Hence for $r < 1$ we find $|\mu|(A) = 0$ and we conclude that $|\mu|(\{|h| < 1\}) = 0$. Combining this with the previous result we get $|\mu|(\{|h| \neq 1\}) = 0$. The function that we look for, is $h 1_{\{|h|=1\}} + 1_{\{|h| \neq 1\}}$. □

Corollary 6.2 *Let μ be a real measure, $h = \frac{d\mu}{d|\mu|}$. Then for any $E \in \Sigma$ we have $\mu^+(E) = |\mu|(1_{E \cap \{h=1\}})$ and $\mu^-(E) = |\mu|(1_{E \cap \{h=-1\}})$ and $\mu^+ \perp \mu^-$. Moreover, if $\mu = \mu_1 - \mu_2$ with positive measures μ_1, μ_2, then $\mu_1 \leq \mu^+$ and $\mu_2 \leq \mu^-$. In this sense the Jordan decomposition is minimal.*

Proof. The representation of μ^+ and μ^- follows from the previous proposition. Minimality is proved as follows. Since $\mu \leq \mu_1$, we have $\mu^+(E) = \mu(E \cap \{h = 1\}) \leq \mu_1(E \cap \{h = 1\}) \leq \mu_1(E)$. □

Proposition 6.3 *If μ is a positive measure and λ a complex measure such that $\lambda \ll \mu$, then $|\lambda| \ll \mu$ and*

$$\frac{d|\lambda|}{d\mu} = |\frac{d\lambda}{d\mu}|.$$

Proof. Exercise 7.8. □

7 Exercises

7.1 Let μ be a real measure on a space (S, Σ). Define $\nu : \Sigma \to [0, \infty)$ by $\nu(E) = \sup\{\mu(F) : F \in \Sigma, F \subset E, \mu(F) \geq 0\}$. Show that ν is a finite positive measure. Give a characterization of ν.

7.2 Prove proposition 4.1.

7.3 Prove a version of proposition 4.3 adapted to the case where $h \in \mathcal{L}^1(S, \Sigma, \mu)$ is complex valued.

7.4 Let X be a symmetric Bernoulli distributed random variable ($\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$) and Y uniformly distributed on $[0, \theta]$ (for some arbitrary $\theta > 0$). Assume that X and Y are independent. Show that the laws $\mathcal{L}_\theta$ ($\theta > 0$) of XY are not absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}$. Find a fixed dominating σ-finite measure μ such that $\mathcal{L}_\theta \ll \mu$ for all θ and determine the corresponding Radon-Nikodym derivatives.

7.5 Let $X_1, X_2, \ldots$ be an independent sequence of symmetric Bernoulli random variables, defined on some probability space. Let

$$X = \sum_{k=1}^{\infty} 2^{-k} X_k.$$

Find the distribution of X. A completely different situation occurs when we ignore the odd numbered random variables. Let

$$Y = 3 \sum_{k=1}^{\infty} 4^{-k} X_{2k},$$

where the factor 3 only appears for esthetic reasons. Show that the distribution function $F : [0, 1] \to \mathbb{R}$ of Y is constant on $(\frac{1}{4}, \frac{3}{4})$, that $F(1 - x) = 1 - F(x)$ and that it satisfies $F(x) = 2F(x/4)$ for $x < \frac{1}{4}$. Make a sketch of F and show that F is continuous, but not absolutely continuous w.r.t. Lebesgue measure. (Hence there is no Borel measurable function f such that $F(x) = \int_{[0,x]} f(u)\,\mathrm{d}u$, $x \in [0, 1]$).

7.6 Let $f \in \mathcal{L}^1(S, \Sigma, \mu)$ be such that $\mu(1_E f) = 0$ for all $E \in \Sigma$. Show that $\mu(\{f \neq 0\}) = 0$. Conclude that the function h in the Radon-Nikodym theorem has the stated uniqueness property.

7.7 Let μ and ν be positive σ-finite measures and λ an arbitrary measure on a measurable space (S, Σ). Assume that $\lambda \ll \nu$ and $\nu \ll \mu$. Show that $\lambda \ll \mu$ and that

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

7.8 Prove proposition 6.3.

7.9 Let λ and μ be positive σ-finite measures on (S, Σ) with $\lambda \ll \mu$. Let $h = \frac{d\lambda}{d\mu}$. Show that $\lambda(\{h = 0\}) = 0$. Show that $\mu(\{h = 0\}) = 0$ iff $\mu \ll \lambda$. What is $\frac{d\mu}{d\lambda}$ if this happens?

7.10 Let μ and ν be positive σ-finite measures and λ a complex measure on (S, Σ). Assume that $\lambda \ll \mu$ and $\nu \ll \mu$ with Radon-Nikodym derivatives h and k respectively. Let $\lambda = \lambda_a + \lambda_s$ be the Lebesgue decomposition of λ w.r.t. μ. Show that (ν-a.e.)

$$\frac{d\lambda_a}{d\nu} = \frac{h}{k} 1_{\{k>0\}}.$$

7.11 Consider the measurable space $(\Omega, \mathcal{F})$ and a measurable map $X : \Omega \to \mathbb{R}^n$ ($\mathbb{R}^n$ is endowed with the usual Borel σ-algebra $\mathcal{B}^n$). Consider two probability measure $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ and let $P = \mathbb{P}^X$ and $Q = \mathbb{Q}^X$ be the corresponding distributions (laws) on $(\mathbb{R}^n, \mathcal{B}^n)$. Assume that P and Q are both absolutely continuous w.r.t. some σ-finite measure (e.g. Lebesgue measure), with corresponding Radon-Nikodym derivatives (in this context often called densities) f and g respectively, so $f, g : \mathbb{R}^n \to [0, \infty)$. Assume that $g > 0$. Show that for $\mathcal{F} = \sigma(X)$ it holds that $\mathbb{P} \ll \mathbb{Q}$ and that (look at excercise 7.10) the Radon-Nikodym derivative here can be taken as the *likelihood ratio*

$$\omega \mapsto \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega) = \frac{f(X(\omega))}{g(X(\omega))}.$$