Radon–Nikodym theorem

In [mathematics](), the **Radon–Nikodym theorem** is a result in [measure theory]() that expresses the relationship between two measures defined on the same [measurable space](). A *measure* is a [set function]() that assigns a consistent magnitude to the measurable subsets of a measurable space. Examples of a measure include area and volume, where the subsets are sets of points; or the probability of an event, which is a subset of possible outcomes within a wider [probability space]().

One way to derive a new measure from one already given is to assign a density to each point of the space, then [integrate]() over the measurable subset of interest. This can be expressed as

$$\nu(A) = \int_A f \, d\mu,$$

where ν is the new measure being defined for any measurable subset A and the function f is the density at a given point. The integral is with respect to an existing measure μ, which may often be the canonical [Lebesgue measure]() on the [real line]() $\mathbf{R}$ or the n-dimensional [Euclidean space]() $\mathbf{R}^n$ (corresponding to our standard notions of length, area and volume). For example, if f represented mass density and μ was the Lebesgue measure in three-dimensional space $\mathbf{R}^3$, then $\nu(A)$ would equal the total mass in a spatial region A.

The Radon–Nikodym theorem essentially states that, under certain conditions, any measure ν can be expressed in this way with respect to another measure μ on the same space. The function f is then called the **Radon–Nikodym derivative** and is denoted by $\frac{d\nu}{d\mu}$.[1] An important application is in [probability theory](), leading to the [probability density function]() of a [random variable]().

The theorem is named after [Johann Radon](), who proved the theorem for the special case where the underlying space is $\mathbf{R}^n$ in 1913, and for [Otto Nikodym]() who proved the general case in 1930.[2] In 1936 [Hans Freudenthal]() generalized the Radon–Nikodym theorem by proving the [Freudenthal spectral theorem](), a result in [Riesz space]() theory; this contains the Radon–Nikodym theorem as a special case.[3]

A [Banach space]() Y is said to have the [Radon–Nikodym property]() if the generalization of the Radon–Nikodym theorem also holds, *mutatis mutandis*, for functions with values in Y. All [Hilbert spaces]() have the Radon–Nikodym property.

Formal description

Radon–Nikodym theorem

The **Radon–Nikodym theorem** involves a measurable space (X, Σ) on which two σ-finite measures are defined, μ and ν. It states that, if $\nu \ll \mu$ (that is, if ν is absolutely continuous with respect to μ), then there exists a Σ-measurable function $f : X \to [0, \infty)$, such that for any measurable set $A \in \Sigma$,

$$\nu(A) = \int_A f \, d\mu.$$

Radon–Nikodym derivative

The function f satisfying the above equality is *uniquely defined up to a μ-null set*, that is, if g is another function which satisfies the same property, then $f = g$ μ-almost everywhere. The function f is commonly written $\dfrac{d\nu}{d\mu}$ and is called the **Radon–Nikodym derivative**. The choice of notation and the name of the function reflects the fact that the function is analogous to a derivative in calculus in the sense that it describes the rate of change of density of one measure with respect to another (the way the Jacobian determinant is used in multivariable integration).

Extension to signed or complex measures

A similar theorem can be proven for signed and complex measures: namely, that if μ is a nonnegative σ-finite measure, and ν is a finite-valued signed or complex measure such that $\nu \ll \mu$, that is, ν is absolutely continuous with respect to μ, then there is a μ-integrable real- or complex-valued function g on X such that for every measurable set A,

$$\nu(A) = \int_A g \, d\mu.$$

Examples

In the following examples, the set X is the real interval [0,1], and Σ is the Borel sigma-algebra on X.

1. μ is the length measure on X. ν assigns to each subset Y of X, twice the length of Y. Then,
   $\frac{d\nu}{d\mu} = 2$.

2. μ is the length measure on X. ν assigns to each subset Y of X, the number of points from the set {0.1, ..., 0.9} that are contained in Y. Then, ν is not absolutely-continuous with respect

to μ since it assigns non-zero measure to zero-length points. Indeed, there is no derivative $\frac{d\nu}{d\mu}$: there is no finite function that, when integrated e.g. from $(0.1 - \varepsilon)$ to $(0.1 + \varepsilon)$, gives 1 for all $\varepsilon > 0$.

3. $\mu = \nu + \delta_0$, where ν is the length measure on X and δ_0 is the Dirac measure on 0 (it assigns a measure of 1 to any set containing 0 and a measure of 0 to any other set). Then, ν is absolutely continuous with respect to μ, and $\frac{d\nu}{d\mu} = 1_{X\setminus\{0\}}$ – the derivative is 0 at $x = 0$ and 1 at $x > 0$.[4]

Properties

- Let ν, μ, and λ be σ-finite measures on the same measurable space. If $\nu \ll \lambda$ and $\mu \ll \lambda$ (ν and μ are both absolutely continuous with respect to λ), then

$$\frac{d(\nu + \mu)}{d\lambda} = \frac{d\nu}{d\lambda} + \frac{d\mu}{d\lambda} \quad \lambda\text{-almost everywhere.}$$

- If $\nu \ll \mu \ll \lambda$, then

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \quad \lambda\text{-almost everywhere.}$$

- In particular, if $\mu \ll \nu$ and $\nu \ll \mu$, then

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1} \quad \nu\text{-almost everywhere.}$$

- If $\mu \ll \lambda$ and g is a μ-integrable function, then

$$\int_X g\,d\mu = \int_X g\frac{d\mu}{d\lambda}\,d\lambda.$$

- If ν is a finite signed or complex measure, then

$$\frac{d|\nu|}{d\mu} = \left|\frac{d\nu}{d\mu}\right|.$$

Applications

Probability theory

The theorem is very important in extending the ideas of probability theory from probability masses and probability densities defined over real numbers to probability measures defined over arbitrary sets. It tells if and how it is possible to change from one probability measure to another. Specifically, the probability density function of a random variable is the Radon–Nikodym derivative of the induced measure with respect to some base measure (usually the Lebesgue measure for continuous random variables).

For example, it can be used to prove the existence of conditional expectation for probability measures. The latter itself is a key concept in probability theory, as conditional probability is just a special case of it.

Financial mathematics

Amongst other fields, financial mathematics uses the theorem extensively, in particular via the Girsanov theorem. Such changes of probability measure are the cornerstone of the rational pricing of derivatives and are used for converting actual probabilities into those of the risk neutral probabilities.

Information divergences

If μ and ν are measures over X, and $\mu \ll \nu$

- The Kullback–Leibler divergence from ν to μ is defined to be

$$D_{\mathrm{KL}}(\mu \parallel \nu) = \int_X \log\left(\frac{d\mu}{d\nu}\right) \, d\mu.$$

- For $\alpha > 0$, $\alpha \neq 1$ the Rényi divergence of order α from ν to μ is defined to be

$$D_\alpha(\mu \parallel \nu) = \frac{1}{\alpha - 1} \log\left(\int_X \left(\frac{d\mu}{d\nu}\right)^{\alpha-1} d\mu\right).$$

The assumption of σ-finiteness

The Radon–Nikodym theorem above makes the assumption that the measure μ with respect to which one computes the rate of change of ν is σ-finite.

Negative example

Here is an example when μ is not σ-finite and the Radon–Nikodym theorem fails to hold.

Consider the Borel σ-algebra on the real line. Let the counting measure, μ, of a Borel set A be defined as the number of elements of A if A is finite, and ∞ otherwise. One can check that μ is indeed a measure. It is not σ-finite, as not every Borel set is at most a countable union of finite sets. Let ν be the usual Lebesgue measure on this Borel algebra. Then, ν is absolutely continuous with respect to μ, since for a set A one has $\mu(A) = 0$ only if A is the empty set, and then $\nu(A)$ is also zero.

Assume that the Radon–Nikodym theorem holds, that is, for some measurable function f one has

$$\nu(A) = \int_A f \, d\mu$$

for all Borel sets. Taking A to be a singleton set, $A = \{a\}$, and using the above equality, one finds

$$0 = f(a)$$

for all real numbers a. This implies that the function f, and therefore the Lebesgue measure ν, is zero, which is a contradiction.

Positive result

Assuming $\nu \ll \mu$, the Radon–Nikodym theorem also holds if μ is localizable and ν is *accessible with respect to μ,*[5]:p. 189, Exercise 9O i.e., $\nu(A) = \sup\{\nu(B) : B \in \mathcal{P}(A) \cap \mu^{\mathbf{pre}}(\mathbb{R}_{\geq 0})\}$ for all $A \in \Sigma$.[6]:Theorem 1.111 (Radon–Nikodym, II) [5]:p. 190, Exercise 9T(ii)

Proof

This section gives a measure-theoretic proof of the theorem. There is also a functional-analytic proof, using Hilbert space methods, that was first given by von Neumann.

For finite measures μ and ν, the idea is to consider functions f with $f \, d\mu \leq d\nu$. The supremum of all such functions, along with the monotone convergence theorem, then furnishes the Radon–Nikodym derivative. The fact that the remaining part of μ is singular with respect to ν follows from a technical fact about finite measures. Once the result is established for finite measures, extending to σ-finite, signed, and complex measures can be done naturally. The details are given below.

For finite measures

Constructing an extended-valued candidate First, suppose μ and ν are both finite-valued nonnegative measures. Let F be the set of those extended-value measurable functions $f : X \to [0, \infty]$ such that:

$$\forall A \in \Sigma: \qquad \int_A f \, d\mu \leq \nu(A)$$

$F \neq \varnothing$, since it contains at least the zero function. Now let $f_1, f_2 \in F$, and suppose A is an arbitrary measurable set, and define:

$$A_1 = \{x \in A : f_1(x) > f_2(x)\},$$
$$A_2 = \{x \in A : f_2(x) \geq f_1(x)\}.$$

Then one has

$$\int_A \max\{f_1, f_2\} \, d\mu = \int_{A_1} f_1 \, d\mu + \int_{A_2} f_2 \, d\mu \le \nu(A_1) + \nu(A_2) = \nu(A),$$

and therefore, $\max\{f_1, f_2\} \in F$.

Now, let $\{f_n\}$ be a sequence of functions in F such that

$$\lim_{n \to \infty} \int_X f_n \, d\mu = \sup_{f \in F} \int_X f \, d\mu.$$

By replacing f_n with the maximum of the first n functions, one can assume that the sequence $\{f_n\}$ is increasing. Let g be an extended-valued function defined as

$$g(x) := \lim_{n \to \infty} f_n(x).$$

By Lebesgue's monotone convergence theorem, one has

$$\lim_{n \to \infty} \int_A f_n \, d\mu = \int_A \lim_{n \to \infty} f_n(x) \, d\mu(x) = \int_A g \, d\mu \le \nu(A)$$

for each $A \in \Sigma$, and hence, $g \in F$. Also, by the construction of g,

$$\int_X g \, d\mu = \sup_{f \in F} \int_X f \, d\mu.$$

Proving equality Now, since $g \in F$,

$$\nu_0(A) := \nu(A) - \int_A g \, d\mu$$

defines a nonnegative measure on Σ. To prove equality, we show that $\nu_0 = 0$.

Suppose $\nu_0 \ne 0$; then, since μ is finite, there is an $\varepsilon > 0$ such that $\nu_0(X) > \varepsilon\,\mu(X)$. To derive a contradiction from $\nu_0 \ne 0$, we look for a positive set $P \in \Sigma$ for the signed measure $\nu_0 - \varepsilon\,\mu$ (i.e. a measurable set P, all of whose measurable subsets have non-negative $\nu_0 - \varepsilon\mu$ measure), where also P has positive μ-measure. Conceptually, we're looking for a set P, where $\nu_0 \ge \varepsilon\,\mu$ in every part of P. A convenient approach is to use the Hahn decomposition (P, N) for the signed measure $\nu_0 - \varepsilon\,\mu$.

Note then that for every $A \in \Sigma$ one has $\nu_0(A \cap P) \ge \varepsilon\,\mu(A \cap P)$, and hence,

$$\begin{aligned}
\nu(A) &= \int_A g \, d\mu + \nu_0(A) \\
&\ge \int_A g \, d\mu + \nu_0(A \cap P) \\
&\ge \int_A g \, d\mu + \varepsilon\mu(A \cap P) = \int_A (g + \varepsilon 1_P) \, d\mu,
\end{aligned}$$

where 1_P is the indicator function of P. Also, note that $\mu(P) > 0$ as desired; for if $\mu(P) = 0$, then (since ν is absolutely continuous in relation to μ) $\nu_0(P) \leq \nu(P) = 0$, so $\nu_0(P) = 0$ and

$$\nu_0(X) - \varepsilon\mu(X) = (\nu_0 - \varepsilon\mu)(N) \leq 0,$$

contradicting the fact that $\nu_0(X) > \varepsilon\mu(X)$.

Then, since also

$$\int_X (g + \varepsilon 1_P)\, d\mu \leq \nu(X) < +\infty,$$

$g + \varepsilon\, 1_P \in F$ and satisfies

$$\int_X (g + \varepsilon 1_P)\, d\mu > \int_X g\, d\mu = \sup_{f \in F} \int_X f\, d\mu.$$

This is impossible because it violates the definition of a supremum; therefore, the initial assumption that $\nu_0 \neq 0$ must be false. Hence, $\nu_0 = 0$, as desired.

Restricting to finite values Now, since g is μ-integrable, the set $\{x \in X : g(x) = \infty\}$ is μ-null. Therefore, if a f is defined as

$$f(x) = \begin{cases} g(x) & \text{if } g(x) < \infty \\ 0 & \text{otherwise,} \end{cases}$$

then f has the desired properties.

Uniqueness As for the uniqueness, let $f, g : X \to [0, \infty)$ be measurable functions satisfying

$$\nu(A) = \int_A f\, d\mu = \int_A g\, d\mu$$

for every measurable set A. Then, $g - f$ is μ-integrable, and

$$\int_A (g - f)\, d\mu = 0.$$

In particular, for $A = \{x \in X : f(x) > g(x)\}$, or $\{x \in X : f(x) < g(x)\}$. It follows that

$$\int_X (g - f)^+\, d\mu = 0 = \int_X (g - f)^-\, d\mu,$$

and so, that $(g - f)^+ = 0$ μ-almost everywhere; the same is true for $(g - f)^-$, and thus, $f = g$ μ-almost everywhere, as desired.

For σ-finite positive measures

If μ and ν are σ-finite, then X can be written as the union of a sequence $\{B_n\}_n$ of disjoint sets in Σ, each of which has finite measure under both μ and ν. For each n, by the finite case, there is a Σ-

measurable function $f_n : B_n \to [0, \infty)$ such that

$$\nu_n(A) = \int_A f_n \, d\mu$$

for each Σ-measurable subset A of B_n. The sum $\left(\sum_n f_n \mathbf{1}_{B_n} \right) := f$ of those functions is then the required function such that $\nu(A) = \int_A f \, d\mu$.

As for the uniqueness, since each of the f_n is μ-almost everywhere unique, so is f.

For signed and complex measures

If ν is a σ-finite signed measure, then it can be Hahn–Jordan decomposed as $\nu = \nu^+ - \nu^-$ where one of the measures is finite. Applying the previous result to those two measures, one obtains two functions, $g, h : X \to [0, \infty)$, satisfying the Radon–Nikodym theorem for ν^+ and ν^- respectively, at least one of which is μ-integrable (i.e., its integral with respect to μ is finite). It is clear then that $f = g - h$ satisfies the required properties, including uniqueness, since both g and h are unique up to μ-almost everywhere equality.

If ν is a complex measure, it can be decomposed as $\nu = \nu_1 + i\nu_2$, where both ν_1 and ν_2 are finite-valued signed measures. Applying the above argument, one obtains two functions, $g, h : X \to [0, \infty)$, satisfying the required properties for ν_1 and ν_2, respectively. Clearly, $f = g + ih$ is the required function.

The Lebesgue decomposition theorem

Lebesgue's decomposition theorem shows that the assumptions of the Radon–Nikodym theorem can be found even in a situation which is seemingly more general. Consider a σ-finite positive measure μ on the measure space (X, Σ) and a σ-finite signed measure ν on Σ, without assuming any absolute continuity. Then there exist unique signed measures ν_a and ν_s on Σ such that $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$, and $\nu_s \perp \mu$. The Radon–Nikodym theorem can then be applied to the pair ν_a, μ.

See also

- Girsanov theorem

- Radon–Nikodym set

Notes

1. Billingsley, Patrick (1995). *Probability and Measure* (Third ed.). New York: John Wiley & Sons. pp. 419–427. ISBN 0-471-00710-2.

2. Nikodym, O. (1930). "Sur une généralisation des intégrales de M. J. Radon" (http://matwbn.ic m.edu.pl/ksiazki/fm/fm15/fm15114.pdf) (PDF). *Fundamenta Mathematicae* (in French). **15**: 131–179. doi:10.4064/fm-15-1-131-179 (https://doi.org/10.4064%2Ffm-15-1-131-179). JFM 56.0922.02 (https://zbmath.org/?format=complete&q=an:56.0922.02). Retrieved 2018-01-30.

3. Zaanen, Adriaan C. (1996). *Introduction to Operator Theory in Riesz Spaces*. Springer. ISBN 3-540-61989-5.

4. "Calculating Radon Nikodym derivative" (https://math.stackexchange.com/q/588442/2978 0). *Stack Exchange*. April 7, 2018.

5. Brown, Arlen; Pearcy, Carl (1977). *Introduction to Operator Theory I: Elements of Functional Analysis*. ISBN 978-1461299288.

6. Fonseca, Irene; Leoni, Giovanni. *Modern Methods in the Calculus of Variations: L^p Spaces*. Springer. p. 68. ISBN 978-0-387-35784-3.

References

- Lang, Serge (1969). *Analysis II: Real analysis*. Addison-Wesley. Contains a proof for vector measures assuming values in a Banach space.

- Royden, H. L.; Fitzpatrick, P. M. (2010). *Real Analysis* (4th ed.). Pearson. Contains a lucid proof in case the measure *v* is not σ-finite.

- Shilov, G. E.; Gurevich, B. L. (1978). *Integral, Measure, and Derivative: A Unified Approach*. Richard A. Silverman, trans. Dover Publications. ISBN 0-486-63519-8.

- Stein, Elias M.; Shakarchi, Rami (2005). *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton lectures in analysis. Princeton, N.J: Princeton University Press. ISBN 978-0-691-11386-9. Contains a proof of the generalisation.

- Teschl, Gerald. "Topics in Real and Functional Analysis" (https://www.mat.univie.ac.at/~gerald/f tp/book-fa/index.html). (lecture notes).