Computational Statistics & Machine Learning

Lecture 4

Lebesgue Integral and Measure

Mark Girolami
mag92@cam.ac.uk

Department of Engineering
University of Cambridge

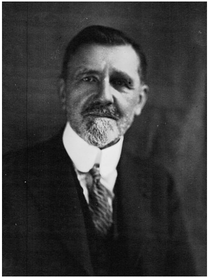
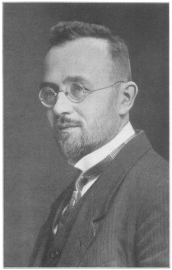
September 16, 2021

Overview

- Definition of Measure of Sets.
- Lebesgue Measure.
- Defining Probability Measures.
- Random Variables
- Radon-Nikodym Derivative

Definition of Measure of Sets

- Theory developed, by amongst others, Henri Lebesgue
- Generalisation of integration to more general spaces than $\mathbb{R}^d$
- Provided means to define axiomatically probability theory

- Emile Borel 1932, Henri Lebesgue, Johan Radon 1920,

Definition of Measure of Sets

- Conceptually all we are doing is generalising the notion of *length* on $\mathbb{R}$, *area* in $\mathbb{R}^2$, and *volume* in $\mathbb{R}^3$, as we met in the definition of the Lebesgue integral
- Consider a set of objects X.
- A set Σ is generated from the original set of objects X which contains the set X, the empty set, and other subsets of X, and is closed under *countable* complementation, union and intersections
- The infinite set of integers $\mathbb{Z}$ is countable, any set that can be put into one-to-one correspondence with the elements of $\mathbb{Z}$ is said to be *countable*
- This set has the somewhat exotic title of a sigma-algebra - as it is stable under union or summation

Definition of Measure of Sets

- For all sets E that belong to Σ, $E \in \Sigma$ then $\mu(E) \geq 0$, so all elements of Σ are measureable
- Obviously the empty set $\mu(\emptyset) = 0$ and if $E_i \subseteq E_j$ then $\mu(E_i) \leq \mu(E_j)$
- The measure assigned to the (countable) union of disjoint sets in Σ is the sum of the measures of each set

$$\mu\left(\cup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$$

- Both sets X, Σ, i.e. $\{X, \Sigma\}$ defines a *Measurable Set*, and along with the measure μ i.e. $\{X, \Sigma, \mu\}$ defines what is called a *Measure Space*
- If (X, Σ_X) and (Y, Σ_Y) are both measurable spaces then a function $f : X \to Y$ is called a *measurable function* if for every measurable set under Y i.e. $B \in \Sigma_Y$ the pre-image $f^{-1}(B)$ is X-measurable i.e. $f^{-1}(B) \in \Sigma_X$

Lebesgue Measure.

- The construction of Lebesgue Measure is highly technical - whole courses devoted to this measure
- A continuous line segment on $\mathbb{R}$, say $I = [a, b]$ has Lebesgue measure which is simply the length of the segment $\mu(I) = b - a$
- There are many technical issues, for example the set of rational numbers $\mathbb{Q}$ form part of the real line
- The Lebesgue measure of a single rational number, say $\frac{1}{4}$, is $\mu(\frac{1}{4}) = 0$, now by additivity of measures $\mu(\mathbb{Q}) = 0$
- Countable sets have Lebesgue measure zero

Defining Probability Measures

- Formal definition of measure theoretic foundations of Probability Theory

- Andrei Kolmogorov 1903 - 1987

Defining Probability Measures

- We have a measure space $\{X, \Sigma, \mu\}$ with the properties of the σ-algebra and measure as before
- One additional property, that is $\mu(X) = 1$ turns the general definition of the measure space into that of a *Probability Space*

Defining Probability Measures

- Defining a set Ω consisting of elements which represents all possible outcomes of an execution of the system, e.g. the single roll of a casino dice would have $\Omega = \{1, 2, 3, 4, 5, 6\}$. The set Ω is typically referred to as the *Sample Space*
- The σ-algebra, usually denoted as $\mathcal{F}$, will be defined in terms of all the outcomes we want to consider, which will be certain subsets of the Sample Space, e.g. the full power set or $\mathcal{F} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ - two outcomes either an even or odd number
- The measure function, the probability is positive and between zero and one, P. The outcomes in $\mathcal{F}$ would each have probability measure $P(\emptyset) = 0, P(\{1, 3, 5\}) = P(\{2, 4, 6\}) = 0.5, P(\Omega) = 1.$
- The probability space is then denoted as $\{\Omega, \mathcal{F}, P\}$

Defining Probability Measures

- For continuous spaces the construction of a measure space is considerably more technical
- Lebesgue measure and the definition of Lebesgue measurable spaces is an enormous theoretical endevour
- We will content ourselves that such spaces exist and can be defined, for exactly the unit line $[0, 1]$ with uniform probability is developed via Lebesgue measure
- We have met the idea of a measurable function already, lets see how this can be useful in an additional definition

Random Variables

- Consider the probability space $\{\Omega, \mathcal{F}, P\}$ and the measurable space $\{E, \mathcal{E}\}$
- As both spaces are measurable there will be an associated measurable function $X : \Omega \rightarrow E$
- From the previous definition the function is measurable if $B \in \mathcal{E}$ the pre-image $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$
- An element of $\mathcal{F}$ is a set of possible outcomes that are measurable and the probability of these outcomes is defined by P
- If E is the real line, the σ-algebra is technically defined by a Borel σ-algebra (long story)
- In this case the function $X : \Omega \rightarrow \mathbb{R}$ is a real-valued *Random Variable* such that $\{\omega : X(\omega) \leq r\} \in \mathcal{F} \;\; \forall \, r \in \mathbb{R}$

Random Variables

- Now we can see the power of the generality of the Lebesgue integral $\int f d\mu$
- Let us define the expectation of a random variable (a measurable function) as defined in the last slide

$$\mathbb{E}\{X\} = \int_{\Omega} X(\omega) dP(\omega)$$

- This is the Lebesgue integral of the measurable function X with respect to the measure (probability) P
- A very general tool for integration as the probability measure and the random function can be defined on any space, e.g. function spaces and stochastic processes

Radon-Nikodym Theorem

- Consider two probability measures, μ and ν on the same measurable space $\{X, \Sigma\}$
- If for every set $A \in \Sigma$ where $\nu(A) = 0$ it holds that $\mu(A) = 0$ then μ is *absolutely continuous* with respect to ν - denoted as $\mu \ll \nu$
- This provides a means to define new measures from old ones - something we utilised - in an ad-hoc manner - when defining the Importance Sampling method
- If P is a probability measure, then a new probability measure Q can be defined if there exists a function $L(\omega)$ such that $dQ(\omega) = L(\omega)dP(\omega)$

Radon-Nikodym Theorem

- For any measurable set B then $Q(B) = \int_B L(\omega) dP(\omega)$
- For this to hold then both $L(\omega) \geq 0$ almost surely with respect to P and $\int_\Omega L(\omega) dP(\omega) = 1$
- In words Q is a probability measure derived from P and $Q(B)$ is the probability of the event under Q.
- Which leads to the relationship you have already exploited

$$\int_\Omega F(\omega) dQ(\omega) = \int_\Omega F(\omega) L(\omega) dP(\omega)$$

- The function

$$L(\omega) = \frac{dQ(\omega)}{dP(\omega)}$$

  is the Radon-Nikodym derivative if $Q \ll P$
- Hugely powerful as it enables definition of probabilities over arbitrary sets e.g. stochastic processes and functions

Radon-Nikodym Theorem

- Finally an example. Consider random variable X on $\mathbb{R}$
- If B is a small neighbourhood around an outcome $X = x \in \mathbb{R}$ then its probability is

$$\int_B dP = \int_B L(x) d\mu(x) = \int_B L(x) dx$$

where dx is Lebesgue Measure on $\mathbb{R}$
- As previously defined

$$L(x) = \frac{dP(x)}{dx}$$

which defines the change of measure from Lebesgue to probability measure P
- Such a Radon-Nikodym derivative w.r.t Lebesgue measure is termed the Probability Density function of P w.r.t. Lebesgue Measure

Radon-Nikodym Theorem

- For a random variable with standard Gaussian measure then

$$dP = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx$$

- Then using Lebesgue integration

$$\begin{aligned} P(X(\omega) \in B) &= \int_\Omega I_{\{X(\omega)\in B\}} dP = \int_B \frac{dP}{dx} dx \\ &= \int_B \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \end{aligned}$$

- Which are the forms of integrals we met in Lecture.1 and Lecture.2, now defined in a systematic and general manner as we shall see in subsequent lectures