Computational Statistics & Machine Learning

Lecture 3

Lebesgue Integral and Measure

Mark Girolami
mag92@cam.ac.uk

Department of Engineering
University of Cambridge

October 19, 2021

Overview

- Riemann Integral.
- Convergence of Functions.
- Lebesgue Integral.

Riemann Integral

- Formal definition and construction of Integral
- Integration is the fundamental operator that appears repeatedly in Machine Learning - it is important

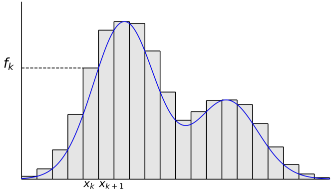
- Bernhard Riemann 1862

Riemann Integral

- We are familiar with integrals of the form

$$I = \int_a^b f(x)dx$$

- The domain ($\mathbb{R}$ or $\mathbb{R}^d$) is split into infinitesimal units
- In the interval $[a, b] \subset \mathbb{R}$ then $\Delta x_k = [x_k, x_{k+1}]$



- So $a = x_1 < x_2 < x_3 < \cdots < x_{n+1} = b$ defines the partition, P, of the domain of integration
- Define two summations

$$S_P(x) = \sum_{k=1}^{n} M_k(x_{k+1} - x_k), \quad s_P(x) = \sum_{k=1}^{n} m_k(x_{k+1} - x_k)$$

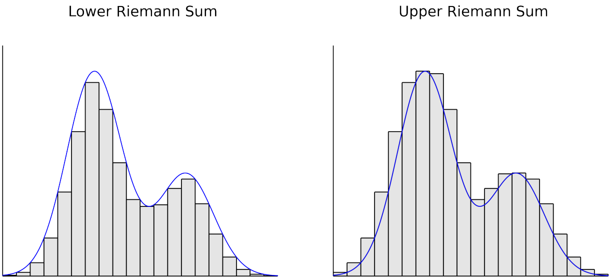
Riemann Integral

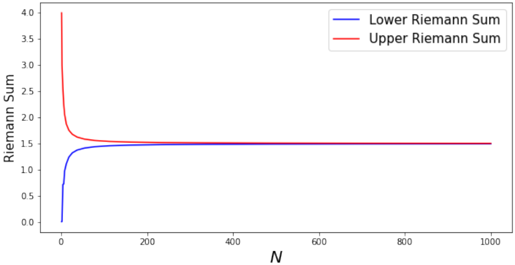
- With $M_k = \sup_{x \in \Delta x_k} f(x)$ and $m_k = \inf_{x \in \Delta x_k} f(x)$ being the *supremum* and *infimum* of $f(x)$ in Δx_k.
- What Riemann does is to take limit as $n \to \infty$ over all possible partitions P to define

$$S(f) = \liminf_{n \to \infty} S_P \quad s(f) = \limsup_{n \to \infty} s_P$$

- Then if $S(f) = s(f) = A$, the value A is the Riemann Integral of $f(x)$ denoted as $A = \int_a^b f(x)dx$
- The bounded function $f(x)$ must be continuous almost everywhere on the domain of integration for the integral to exist

Upper and Lower Riemann Sums





Riemann Integral

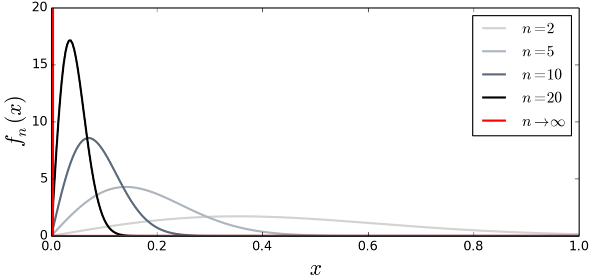
- For applications in Machine Learning the Riemann Integral has a number of shortcomings
- The construction of the integral requires a partition of the domain, for $\mathbb{R}^d$ this is natural.
- What about the domain comprised of continuous functions starting at zero i.e. $C[0, \infty]$, how is this divided into subintervals of equal size?
- Domains other than $\mathbb{R}^d$ are common in Machine Learning and need to be handled appropriately
- The Lebesgue Integral provides an ingenious resolution to this issue

Riemann Integral

- The other shortcoming is that the exchange of limit operations only applies under stringent conditions
- $$\lim_{n\to\infty}\int_{-\infty}^{+\infty} f_n(x)dx = \int_{-\infty}^{+\infty}\lim_{n\to\infty} f_n(x)dx$$
- This is a common operation in Machine Learning applications such as function approximation

Riemann Integral

- Think of the function $f_n(x) = 2n^2 x \exp(-n^2 x^2)$ in the unit interval $[0, 1]$ for all integer n



- The function is continuous on the unit interval and so is Riemann integrable
- $\lim\limits_{n\to\infty} \int_0^1 f_n(x)dx = \lim\limits_{n\to\infty} (1 - \exp(-n^2)) = 1$
- $\int_0^1 \lim\limits_{n\to\infty} f_n(x)dx = \int_0^1 0dx = 0$
- hmmmm what is going on?

Convergence of Functions

- We need to look at the convergence of functions to a limit
- This is important in optimisation for example when training a Deep Net
- There are a number of senses in which a sequence of functions will converge
- The two that we will consider at this point are *Pointwise* and *Uniform* convergence
- We shall observe the difference in these and why the Riemann Integral requires the stricter form of convergence

Pointwise Convergence

- Consider a sequence of real valued functions f_n with $n = 1, 2, \cdots$ i.e. $n \in \mathbb{N}$
- Each f_n is a real valued function $f_n(x)$ of $x \in D \subseteq \mathbb{R}$
- If the sequence $f_1(x), f_2(x), \cdots, f_n(x), \cdots$ converges to $f(x)$ and does so for every $x \in D$
- This is said to converge *pointwise* - at each point x - on D - the pointwise limit is $f(x) = \lim_{n \to \infty} f_n(x)$
- Formally the sequence of functions $f_1(x), f_2(x), \cdots, f_n(x), \cdots$ converges pointwise to f on D if given $\epsilon > 0$ there is a natural number $N = N(\epsilon, x)$ such that for $n > N$ it holds that $|f_n(x) - f(x)| < \epsilon$
- Convergence will vary depending on x in a pointwise manner

Pointwise Convergence

- A simple example illustrates an important facet of pointwise convergence
- The sequence of continuous functions on $[0, 1]$ is defined as the polynomials $f_n(x) = x^n$
- The sequence of continuous functions converges to

$$f(x) = \lim_{n\to\infty} f_n(x) = \begin{cases} 0 & \text{for } \; 0 \leq x < 1 \\ 1 & \text{at } \; x = 1 \end{cases}$$

- This is not a continuous function, the sequence of continuous functions converges to one that is discontinuous at $x = 1$
- Pointwise convergence DOES NOT preserve continuity of functions
- So

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} f_n(x)dx \neq \int_{-\infty}^{+\infty} \lim_{n\to\infty} f_n(x)dx$$

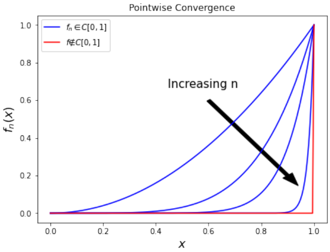
- The Riemann Integral requires a stronger form of convergence of functions that does preserve continuity

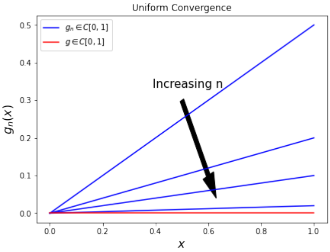
Uniform Convergence

- For each $x \in D$ one can select an $N(\epsilon, x)$ in a pointwise manner so convergence is governed by each point selected
- Alternatively in certain cases one can select $N(\epsilon)$ independent of x so that $|f_n - f| < \epsilon$ for $n > N$ for all $x \in D$. The convergence is uniform over all x
- Formally the sequence of functions $f_1(x), f_2(x), \cdots, f_n(x), \cdots$ converges uniformaly to f on D if given $\epsilon > 0$ there is a natural number $N = N(\epsilon)$ such that for $n > N$ it holds that $|f_n(x) - f(x)| < \epsilon \ \ \forall x \in D$
- Convergence will be uniform and independent of choice of x. Each f_n is contained in an ϵ tube on D around f
- Uniform convergence does preserve continuity as required for the limiting process of Riemann Integration

Pointwise vs. Uniform Convergence

- The function $f_n(x) = x^n$, $x \in [0,1]$ converges pointwise to $f(x) = \begin{cases} 0 & \text{for } x \in [0,1) \\ 1 & \text{for } x = 1 \end{cases}$
- The function $g_n(x) = x/n$, $x \in [0,1]$ converges uniformly to $g(x) = 0$





Lebesgue Integral

- Formal definition and construction of Integral
- Integration is the fundamental operator that appears repeatedly in Machine Learning - it is important

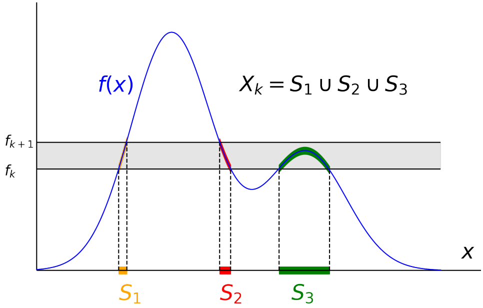
- Henri Lebesgue 1875

Lebesgue Integral

- The construction of the Riemann Integral proceeds by making a partition of the domain
- For the general types of domains we meet in Machine Learning this partitioning is problematic
- What Lebesgue ingeniously did was to partition the range of the function which would take values in $\mathbb{R}$ and so can be done straighforwardly
- Once the range is partitioned the size or measure of parts of the domain that correspond to the range partition are identified
- This exploits the notion of the size or measure of a general set, so for the set of continuous functions a measure needs to be defined - this turns out to be more straightforward than partitioning such domains
- In addition, as the range of the function is partitioned, the strict continuity requirement of the Riemann integral is relaxed

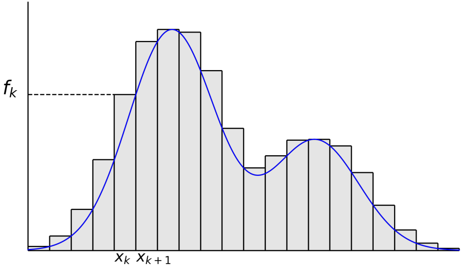
Lebesgue Integral

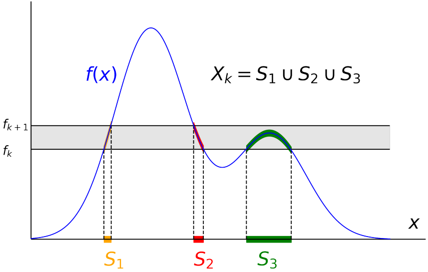
- Consider the bounded function $f(x)$ defined on an abstract set X such that $0 \leq f_{min} \leq f(x) \leq f_{max}$
- Now partition the $f(x)$ axis using the sequence $f_{min} = f_1, f_2, \cdots, f_n = f_{max}$
- There will be sets of values of x such that $f_k \leq f(x) < f_{k+1}$ for $x \in X_k$
- The size (we will now start referring to this as the *measure*) of each of these sets X_k will have value $\mu(X_k)$



Lebesgue Integral

- Compare the Riemann integral to the Lebesgue integral





Lebesgue Integral

- The Lebesgue sum of products of the function value and the measure of the corresponding set is

$$\sum_{k=1}^{n} f_k \mu(X_k)$$

- If the Lebesgue sum is convergent as $n \to \infty$ such that $|f_k - f_{k+1}| \to 0$
- The limiting value is the Lebesgue Integral of the function $f(x)$ over the set X

$$\int_X f d\mu = \lim_{\max |f_k - f_{k-1}| \to 0} \sum_{k=1}^{n} f_k \mu(X_k)$$

Lebesgue Integral

- The Riemann Integral is inextricably linked to the natural ordering of $\mathbb{R}$ - i.e. $1 < 2 < 3.4 < 3.5$ etc... integrating on other structures is challenging
- The Lebesgue integral only requires a measure of sets on the domain of integration so is more general
- Monotone and Dominated convergence Theorems of Lebesgue Integral allow change of order of limit and integration requiring only pointwise convergence
- If a function is Riemann integrable it is Lebesgue integrable - the converse is not strictly the case e.g. discontinuous functions