

Riemann and Lebesgue Integral

New kind of integral used in probability and statistical inference

Riemann Integral

The Riemann integral is a way to define the integral of a function over an interval $[a, b]$ in terms of the sums of areas of rectangles.

A Riemann sum for the function f is given by:

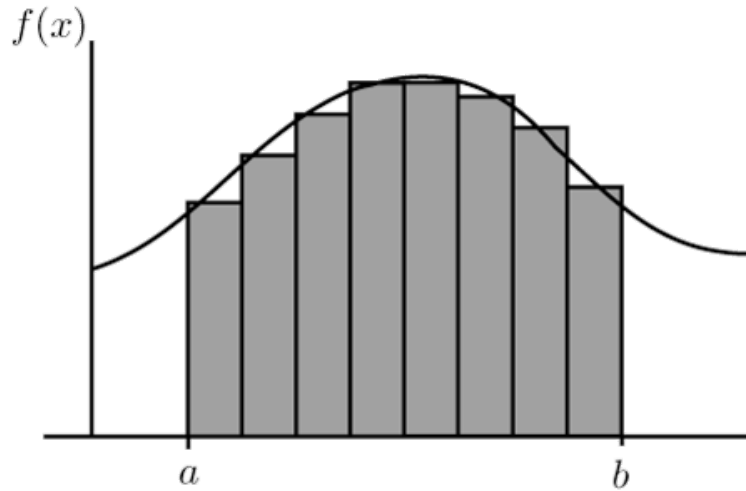
$$S = \sum_{i=1}^n f(x_i^*) \Delta x_i$$

where:

- $[a, b]$ is divided into n subintervals $[x_{i-1}, x_i]$
- $\Delta x_i = x_i - x_{i-1}$ is the width of the i -th subinterval,
- $x_i^* \in [x_{i-1}, x_i]$ is a sample point in the i -th subinterval
- $f(x_i^*) \Delta x_i$ represents the area of a rectangle with height $f(x_i^*)$ and width Δx_i

The Riemann Integral is the limit:

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*) \Delta x_i$$



⊗ Problem

The Riemann integral, while a fundamental concept in mathematical analysis, has some limitations that make it unsuitable for dealing with more general functions, especially when working with functions that are not continuous or have complex discontinuities. In fact, The Riemann integral requires functions to be continuous almost everywhere; if a function has too many discontinuities or they are complexly distributed, the itegral cannot be defined. In addition, the Riemann integral is defined for finite functions over finite intervals. This makes it unsuitable for dealing with unbounded functions or functions defined on infinite sets.

✔ Solution

Per superare queste limitazioni si utilizza un approccio diverso basato sull'integrale di Lebesgue. Tale integrale è il più utilizzato in probabilità.

Funzionamento:

Invece di suddividere il dominio della funzione (come fa l'integrale di Riemann), l'integrale di Lebesgue suddivide il *codominio* della funzione in

intervalli e misura la lunghezza degli insiemi di punti corrispondenti nel dominio (ovvero somma rettangoli che stanno alla stessa altezza).

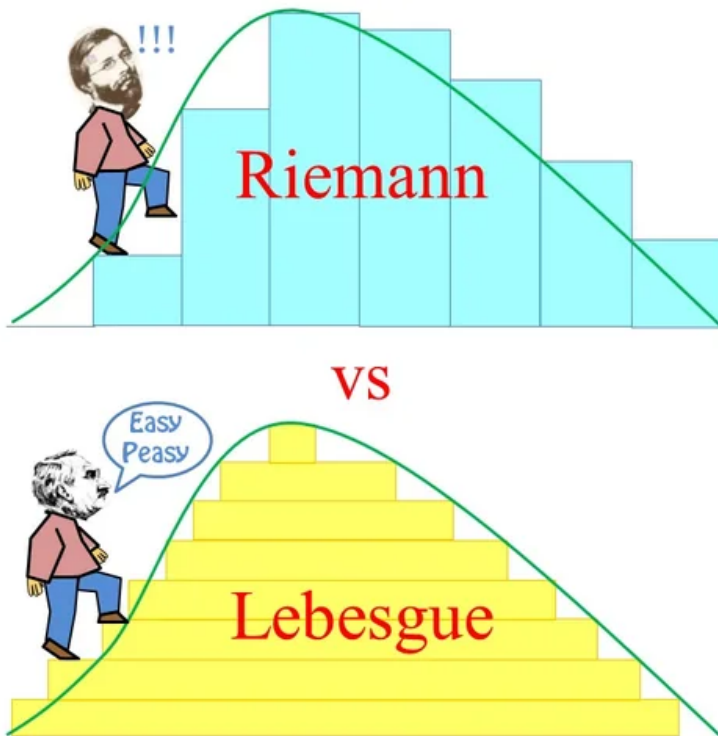
Lebesgue è definito per una classe molto più ampia di funzioni, incluse quelle discontinue su insiemi complessi, può gestire funzioni non limitate e definite su intervalli infiniti, purché la loro misura complessiva sia finita.

Formula Uniformata:

$$\int x d \cdot F(x)$$

where:

- $d \cdot F(x)$ is the function's jump



We know two mechanism to assign probability:

- **Density of probability:** quanto è densa la probabilità di un segmento.
- **Function's jump:** Probabilità presa come salto.

Thanks to Lebesgue's integral, I no longer need the concept of density in statistics because of the unified writing of Lebesgue's integral. Thus, thanks to this new integral we can integrate where Reimann fails.

Now we have the tools to express the mean and variance in the probability theory:

Axiom: $(\Omega, F, P) \rightarrow$ Gormogorov axiom (1933)

- Ω = Universe of possible outcomes
- F Sigma algebra of events
- P Probability measure

Probability is define into the sigma algebra F , infact the probability is a function like this:

$$P : F \rightarrow R$$

Random Variable (Aleatory Variable)

A random variable is a function from Ω in R .

$$X : \Omega \rightarrow R$$

X is similar to the statistics variable

Prendo un outcomes e gli associo un valore, un numero.

Sto dunque creando un nuovo spazio di probabilità dove la nuova Ω sono tutte le misurazioni. Dunque si ottiene (Ω', F', P') .

$X = x_1, x_2, \dots, x_n \rightarrow$ is a set of random variable called **Random sample**

all the x_i are independent each other (for assumption). We can said they are *i. i. d* (identical and independent distributed [they come from the same distribution]).

Random Sample (set of aleatory variable)

$$(X_1, X_2, \dots, X_n)$$

Empirical Sample (all the x_i are mesuration value)

$$x_1, x_2, \dots, x_n$$

Example

Immagina di lanciare un dado 10 volte. Ogni lancio può essere visto come una variabile aleatoria X_i che rappresenta l'esito del dado. Poiché ogni lancio è indipendente e segue la stessa distribuzione (valori uniformemente distribuiti tra 1 e 6), l'insieme dei 10 lanci forma un random sample:

$$(X_1, X_2, \dots, X_{10})$$

Se il risultato dei 10 lanci è:

$$x_1 = 3, x_2 = 5, x_3 = 2, \dots, x_{10} = 6$$

allora x_1, x_2, \dots, x_{10} costituiscono l'empirical sample, cioè i valori concreti osservati.

Mean and Variance, in probability, are aleatory variable.

$$\text{Mean} = \sum_{i=1}^n x_i \cdot P_i \quad \text{Discrete}$$

$$\text{Mean} = \int x f(x) dx \quad \text{Continue}$$

The formula of mean in probability is written in two ways, one for continues function and the other for discrete function. If we use Lebesgue we can use only one for both:

$$\text{Mean} = \int x d \cdot F(x)$$

The **variance** indicates the mean of the deviations from the mean.

$$\text{Variance in Probability} = \int (x - \mu)^2 d \cdot F(x) = \sigma^2$$

Also we can define the Expected value $E(\bar{x}) = \mu$

When i write: $E(x)$ i mean $E(X) = \int x d \cdot F(X)$ (Media del valore atteso)