Measure theory and Integration in STATISTICS

Rajeev pandey

Department of Statistics

DEPARTMENT OF STATISTICS
UNIVERSITY OF LUCKNOW
LUCKNOW-226 007 (INDIA)

PROFRAJEEVLU@GMAIL.COM
PANDEY_RAJEEV@LKOUNIV.AC.IN

Relation between ML and Integration

Statistical Concepts/Techniques	Concepts in Integration Theory
Probability measures	An Integral of some non-negative function w.r.t. a particular measure
Probability of an event	Evaluation of the integral over that event
Mean, Moments, Variance, - --	Integration of a rv/ positive power of a rv,a deviated rv w.r.t. appropriate probability measure **L_p space**

Relation between ML and Integration

Statistical Concepts/Techniques	Concepts in Integration Theory

Covariance, Correlation	Integration of product of two deviated rvs/ two standardized rvs w.r.t. appropriate probability measure

Distributional Convergence In Stochastic Process --	Integral seen as a linear functional on the set of Probability measures

$$P_n \Rightarrow P$$
$$\Leftrightarrow \int_{\Omega} f(w)dP_n \to \int_{\Omega} f(w)dP \quad \forall f \in C(\Omega),$$
$< \Omega, d >$ is a metric space

MEASURE THEORY HELPS STATISTICS IN TWO WAYS

Probability Theory

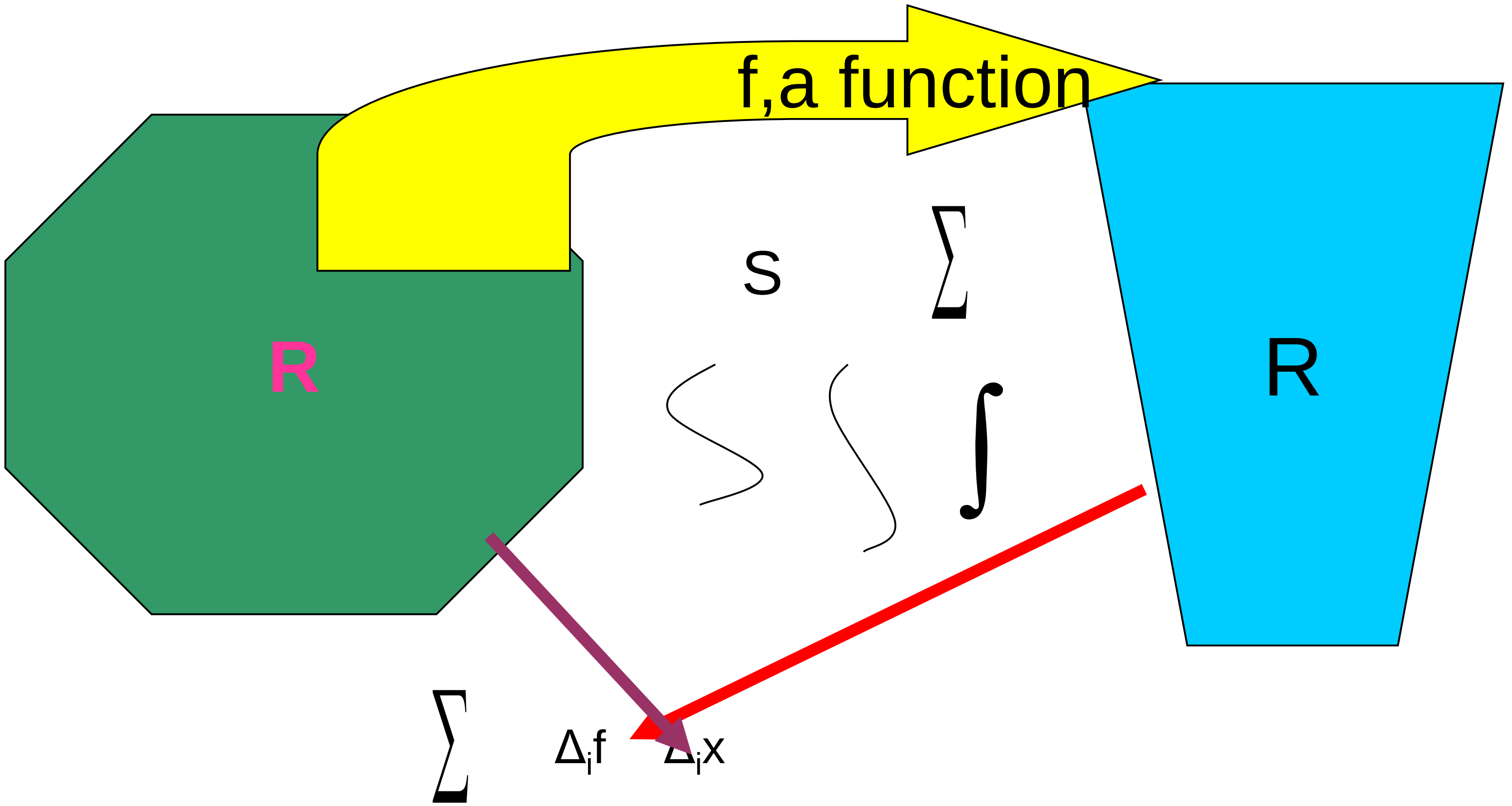
Real Analysis And
Complex Analysis

(Convergence, Limit, Cont., Integration and Diff.)

.Theory of Statistics (Sampling Distribution and Inference)

Analysis and Statistics

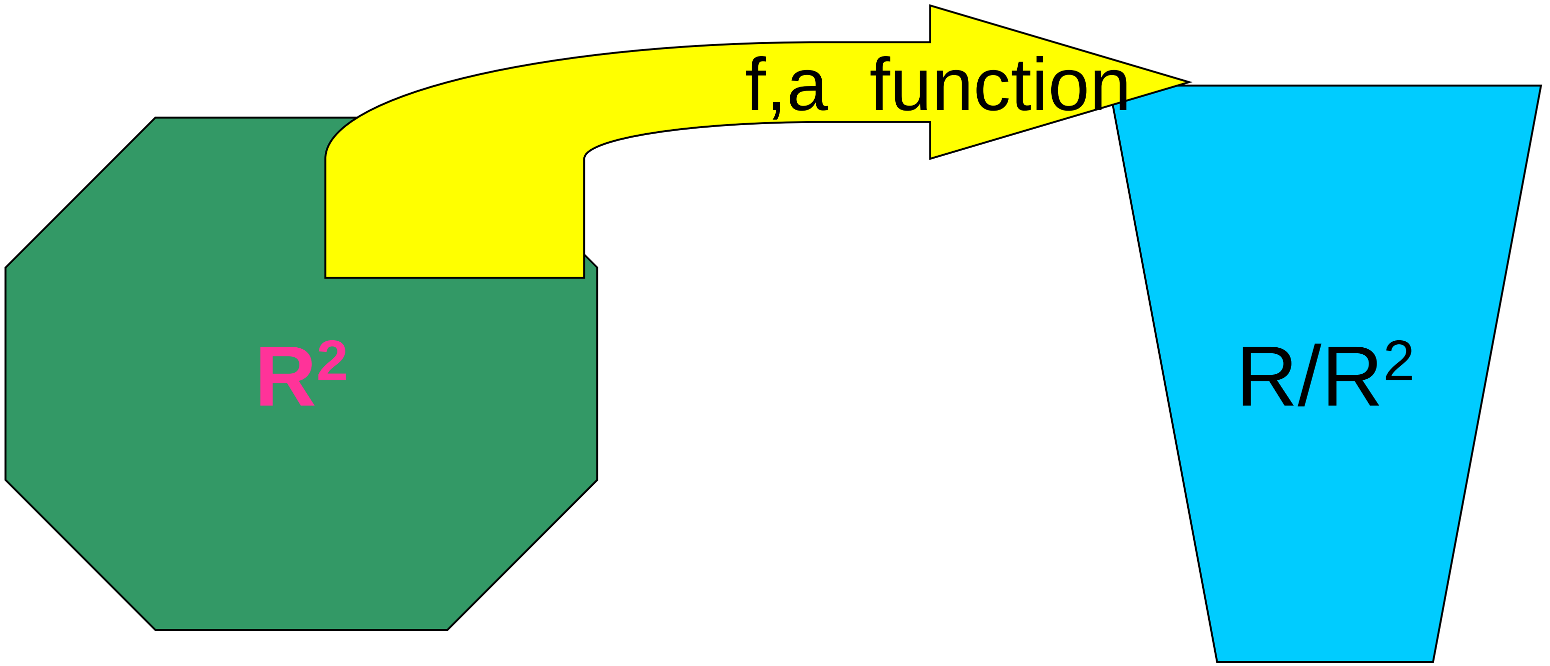
What is Integration?



The product as well as the series should be defined

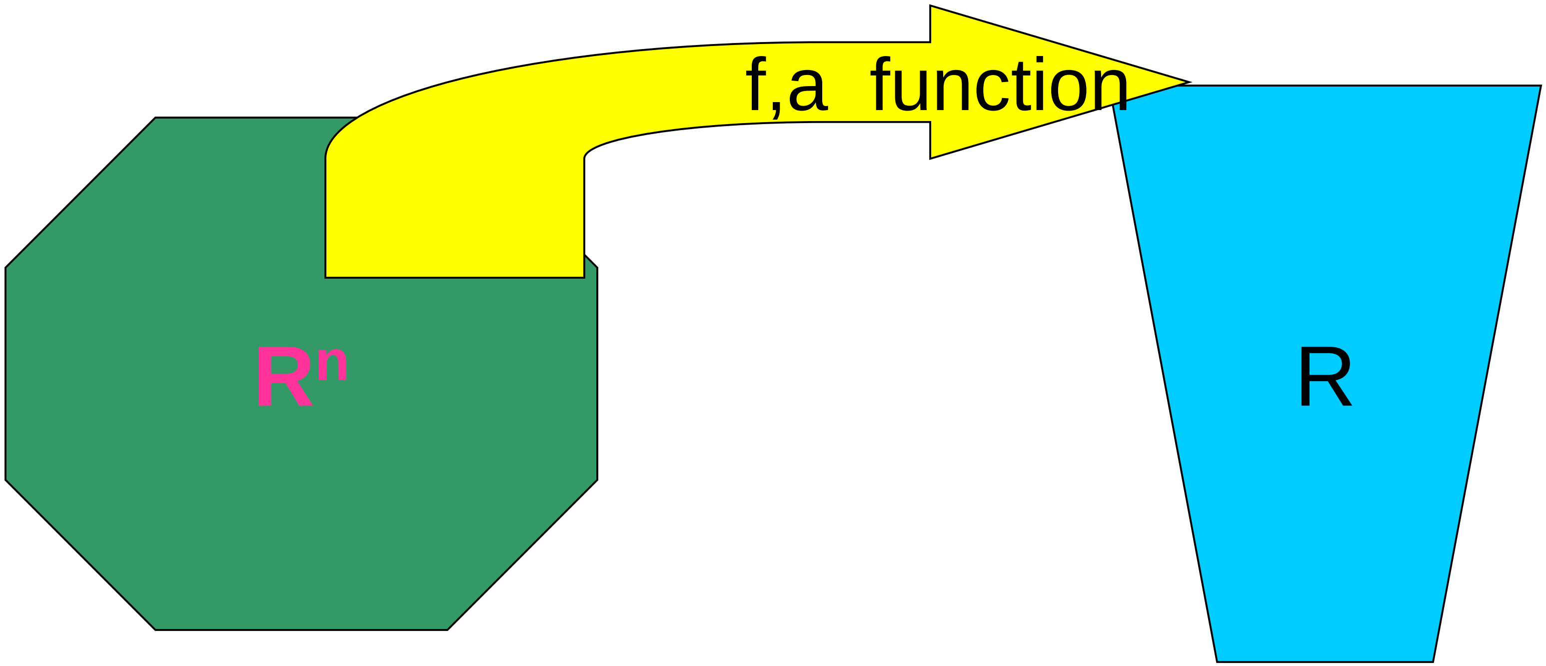
The series should be convergent

What is Integration?



Integration is nothng but infinite sum of infinitesimal quantities

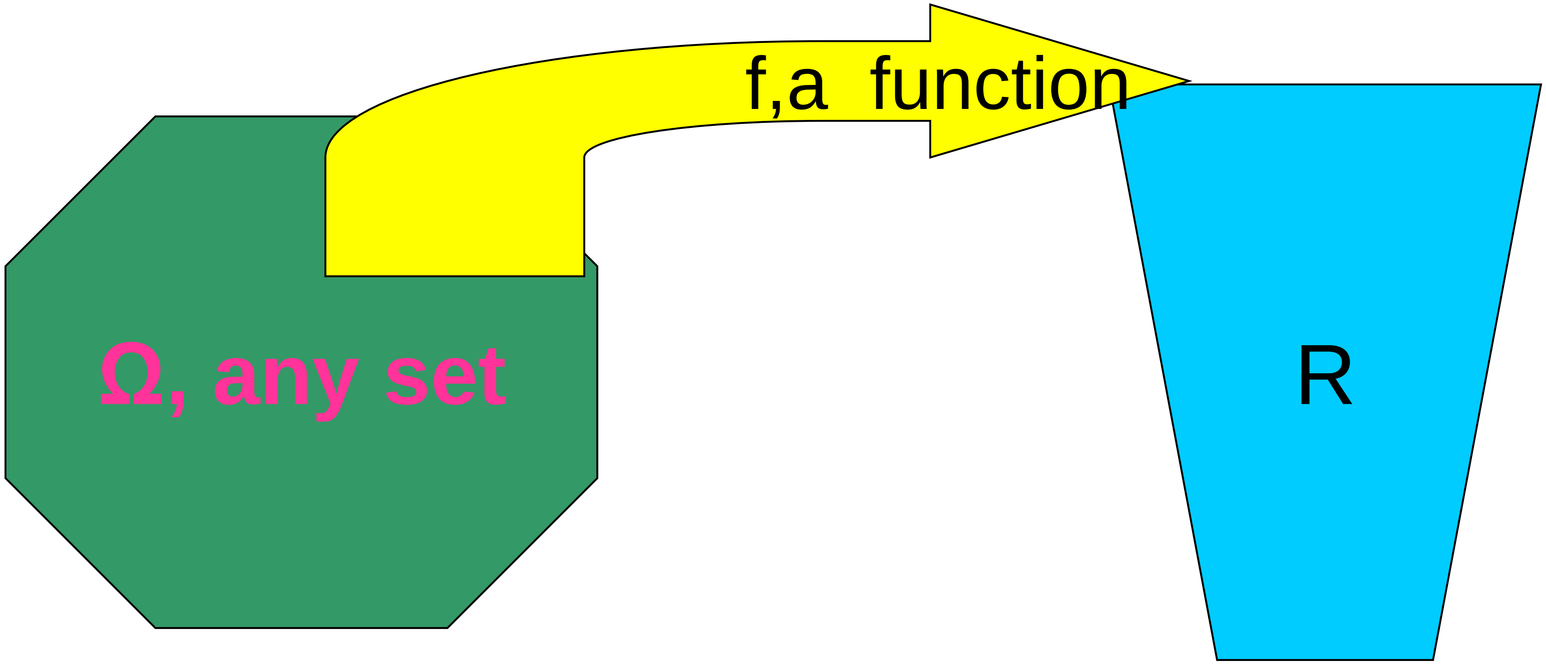
What is Integration?



Integration is nothng but infinite sum of infinitesimal quantities

Whenever we want to measure overall characteristic of something complicated we need some type of integration.

What is Integration?

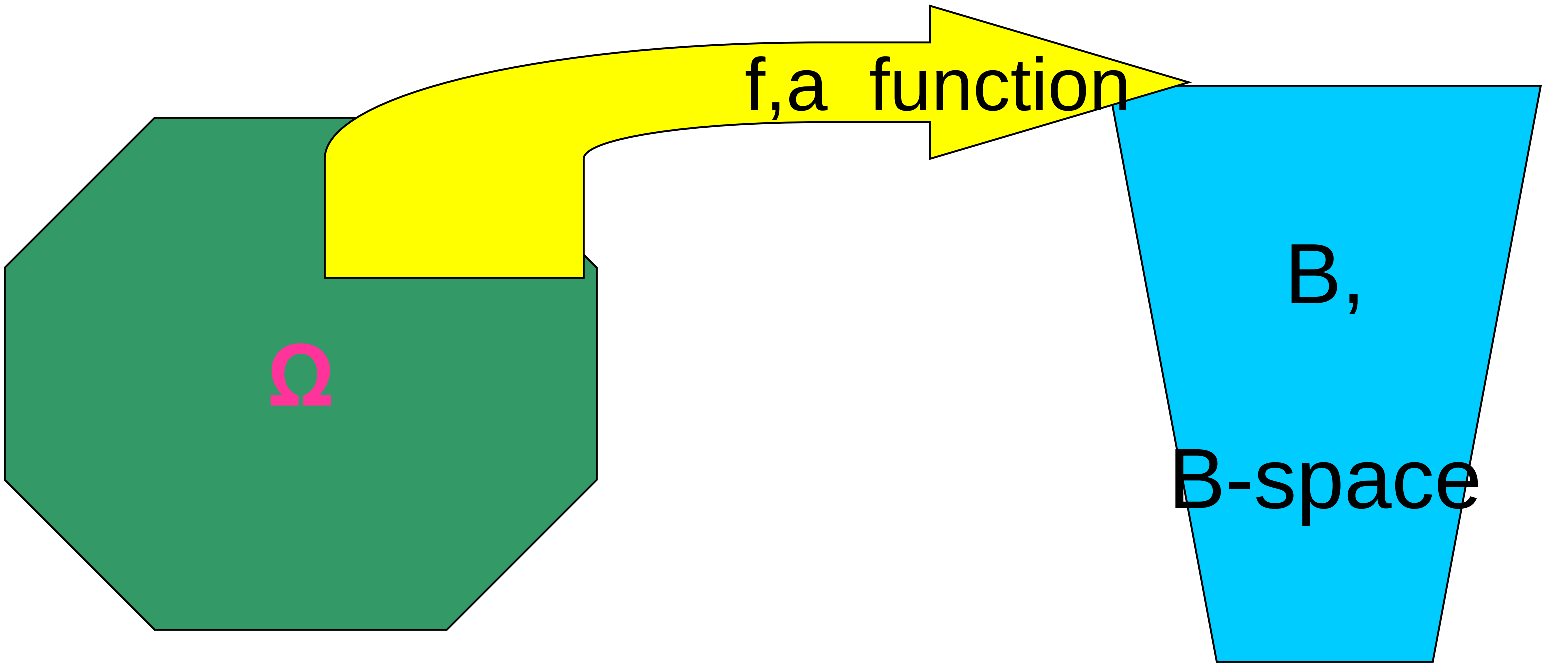


Integration =Integrate+ion

How long will the patient survive?

What percent of people earn more than 50,000 Tk permonth?

What is Integration?



Bochner Integral

History of Integration

Antiphon, Eudoxus, Euclid and Archimedes

Theory of Exhaustion (Archmedes (287BC-202BC)

Riemann Integration(Riemann, G.F.B.(1826-1866))

Stieltjes Integration (Stieltjes,T.J.(1856-1904)

Lebesgue Integration (1902) (Lebesgue,H.(1875-1941)

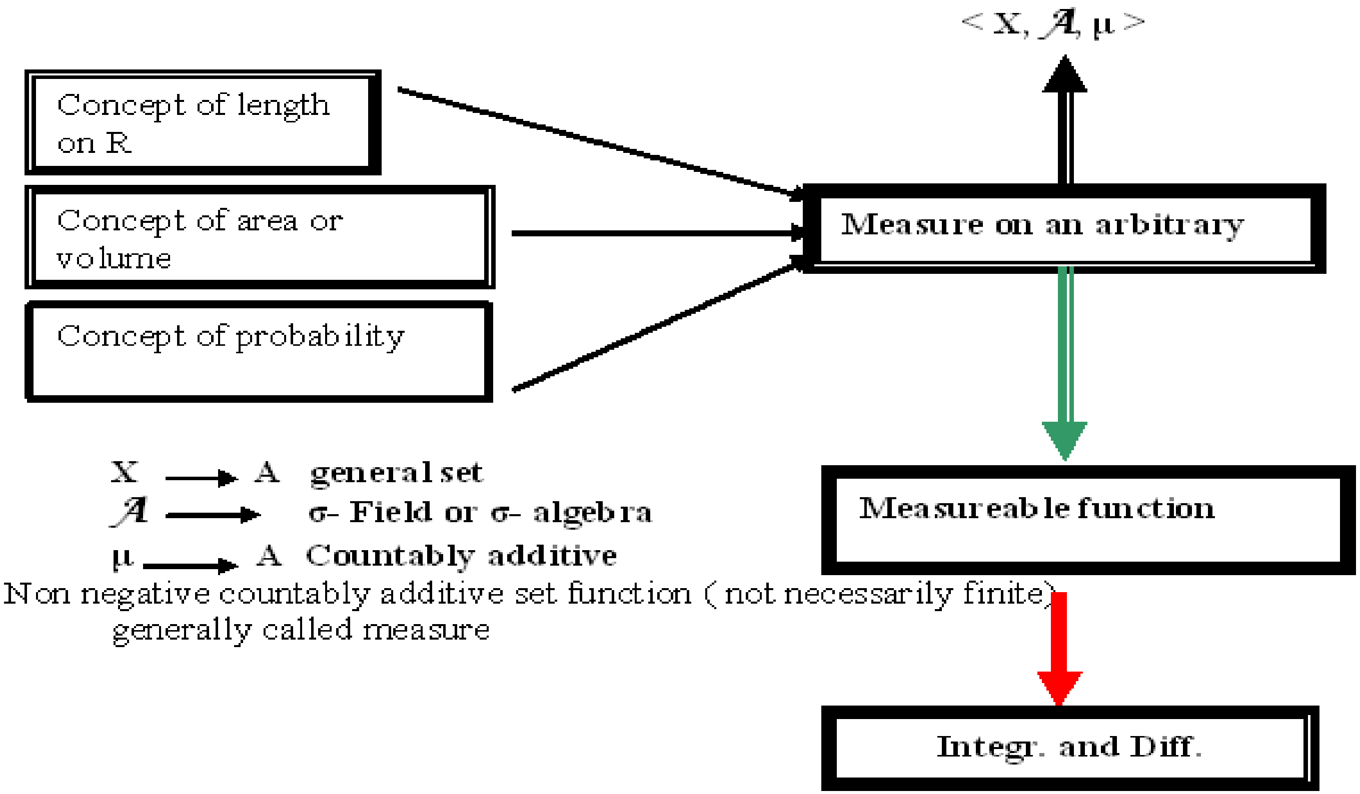
Lebesgue-Stieltjes Integration

General Measure Integration (Bohcner Integral((1933))

Generalization of Integration

GIST OF MEASURE THEORY I

MEASURE AND PROBABILITY MEASURE



$X \longrightarrow$ A **general set**

$\mathcal{A} \longrightarrow$ **σ- Field or σ- algebra**

$\mu \longrightarrow$ A **Countably additive**

Non negative countably additive set function (not necessarily finite) generally called measure

$< X\ \mathcal{A} > \longrightarrow$ Measurable space, A measurable sets

Why σ -field ??

On many important occasions important set function such as area or length can not be defined on every set

Lebesgue versus Riemann

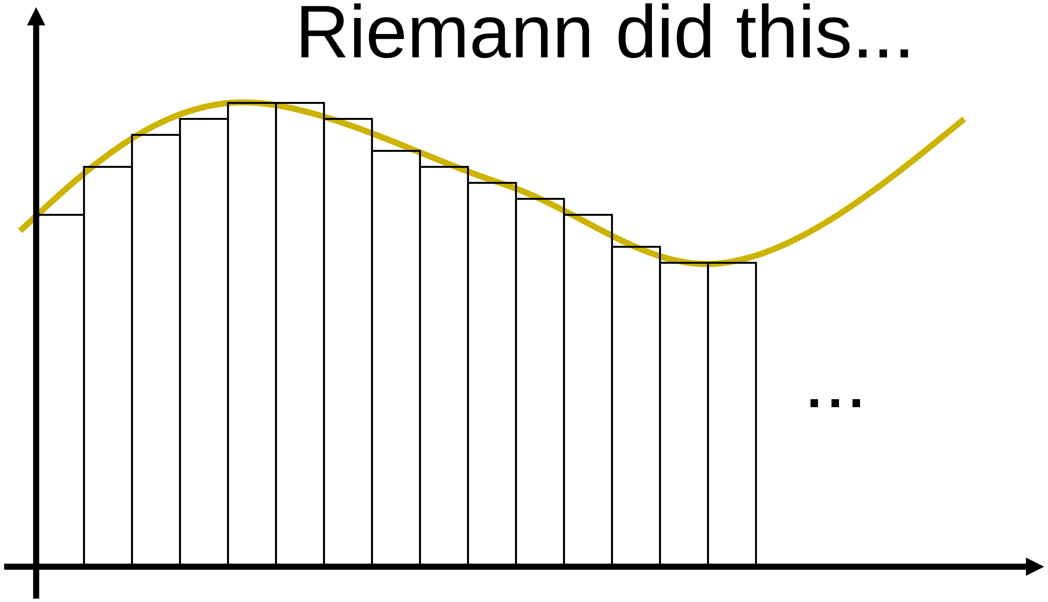
Folland (1984) summarized the difference between the Riemann and Lebesgue approaches thus:

"to compute the Riemann integral of f, one partitions the domain $[a, b]$ into subintervals", while in the Lebesgue integral, "one is in effect partitioning the range of f".

Lebesgue versus Riemann

A little perspective:

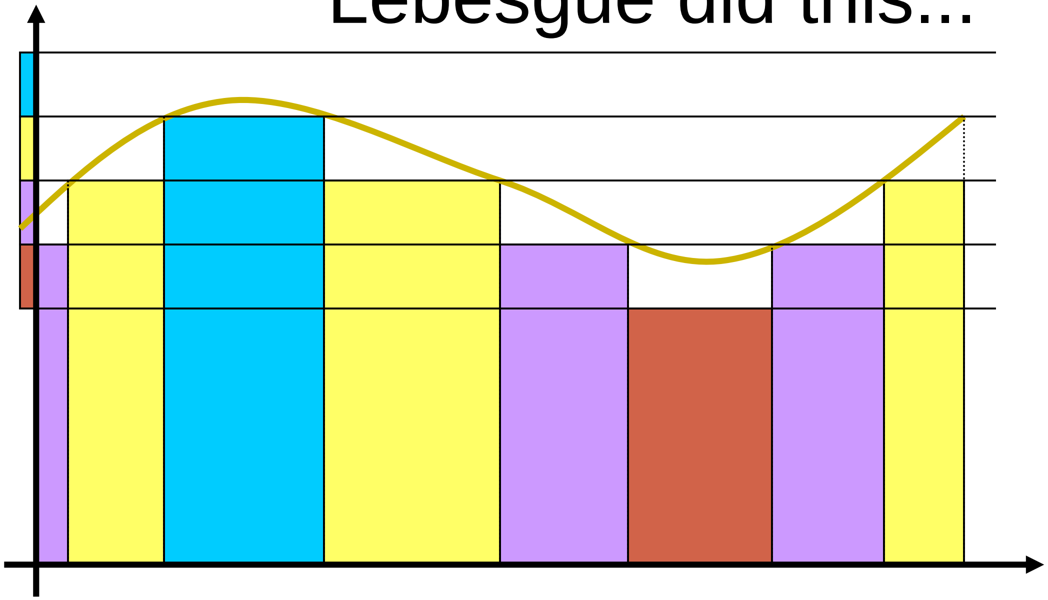
Riemann did this...



Lebesgue versus Riemann

A little perspective:

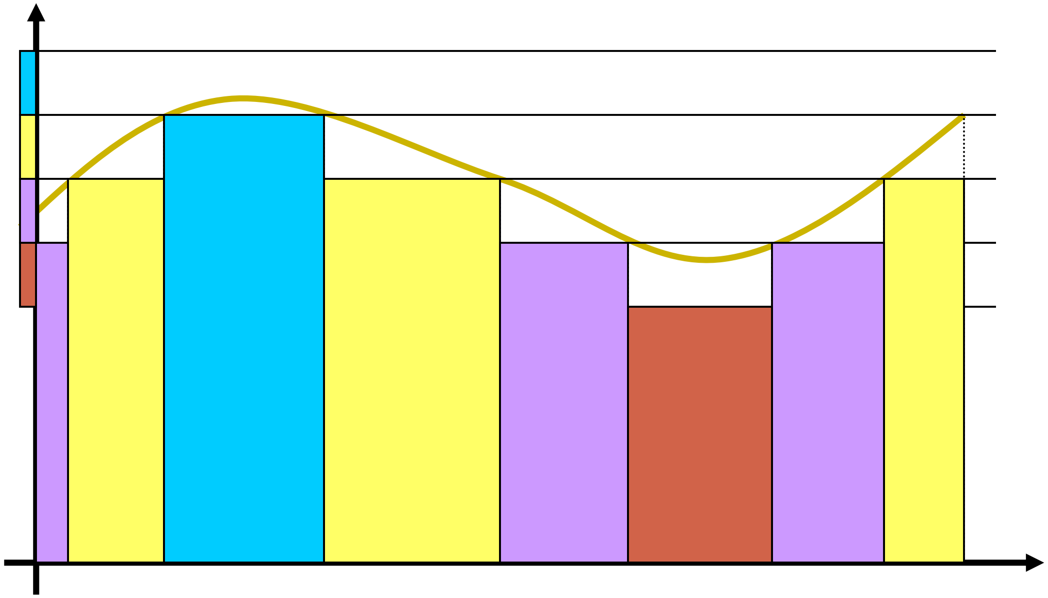
Lebesgue did this...



Lebesgue versus Riemann

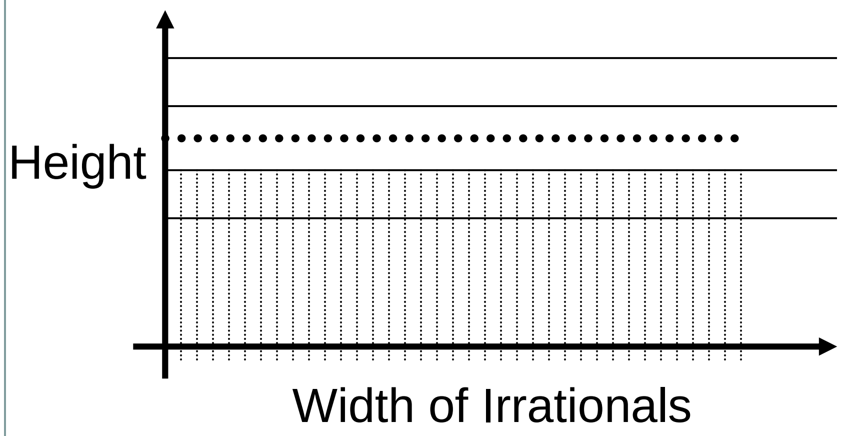
A little perspective:

Lebesgue did this...



Why is this better????

Lebesgue versus Riemann



Consider the function:

$$f = \begin{cases} 1 & \text{at irrationals} \\ 0 & \text{at rationals} \end{cases}$$

If we follow Lebesgue's reasoning, then the integral of this function over the set [0,1] should be:

(Height) x (width of irrationals)
= 1 x (measure of irrationals)

If we can "measure" the size of super level sets, we can integrate a lot of function!

Lebesgue versus Riemann

Measure Theory begins from this simple motivation.

If you remember a couple of simple principles, measure and integration theory becomes quite intuitive.

The first principle is this...

Integration is about functions. Measure theory is about sets. The connection between functions and sets is super/sub-level sets.

Lebesgue versus Riemann

Let f be a bounded function on $[a,b]$:

f is Riemann integrable on $[a,b]$ if and only the set of discontinuities of f *has Lebesgue measure 0*

If f is Riemann integrable on $[a,b]$, f *is Lebesgue integrable on* $[a,b]$.
Two are equal.

Lebesgue versus Riemann

In case of improper integrals

Their does not exist any clear-cut relationship.

f may be not Lebesgue integrable despite its Riemann integrability

The same is true if types of integration change their places

Caution:The type is very important.

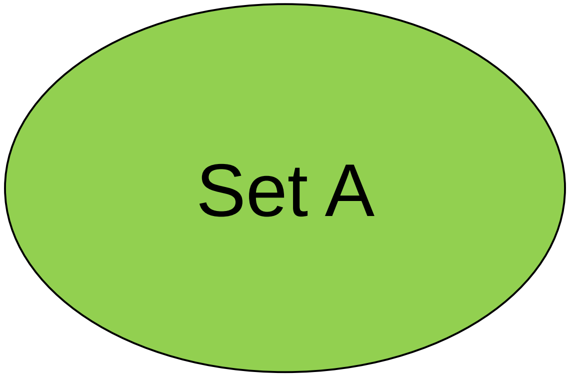
For examples, see “ Lebesgue Measure and Intgration” by Jain and Gupta (1987), pp 165-167.

σ-Algebras??

Measure theory is about measuring the size of things

In math we measure the size of sets.

How big is the set A?



σ-Algebras??

"Size" should have the following property:

The size of the union of disjoint sets should equal the sum of the sizes of the individual sets.

Makes sense.... but there is a problem with this...

Consider the interval [0,1]. It is the disjoint union of all real numbers between 0 and 1. Therefore, according to above, the size of [0,1] should be the sum of the sizes of a single real number.

If the size of a singleton is 0 then the size of [0,1] is 0.
If the size of a singleton is non-zero, then the size of [0,1] is infinity!

This shows we have to be a bit more careful.

Sigma Algebras (σ-algebra)

Consider a set Ω.

Let *F* be a collection of subsets of Ω.

F is a σ-algebra if:

1) $\Omega \in F$

2) $A \in F \Rightarrow A^C \in F$

3) $A_1, A_2 \in F \Rightarrow A_1 \cap A_2 \in F$

4) $A_1, A_2, A_3, \ldots \in F \Rightarrow \bigcup_{i=1}^{\infty} A_i \in F$

An algebra of sets is closed under finite set operations. A σ-algebra is closed under countable set operations. In mathematics, σ often refers to "countable".

General Measurable Functions I

$f: \quad \Omega_1 \Longrightarrow \Omega_2$

Ω_1 and Ω_2 are endowed with σ algebras F_1 and F_2 respectively.

Let us take any B in F_2 and take inverse of B, $f^{-1}\{B\}$. Since B, subset of Ω_2 , $f^{-1}\{B\}$ is a subset $\Omega_{1..}$

If $f^{-1}\{B\}$ belong to F_1 for all B in F_2 , f is called measurable w.r. t. F_1 and F_2

General Measurable Functions II

f: $\Omega_1 \Longrightarrow \Omega_2$

In applications Ω_2 is a metric space and F_{2}, Borel σ-algebra , the σ-algebra generated by open sets.

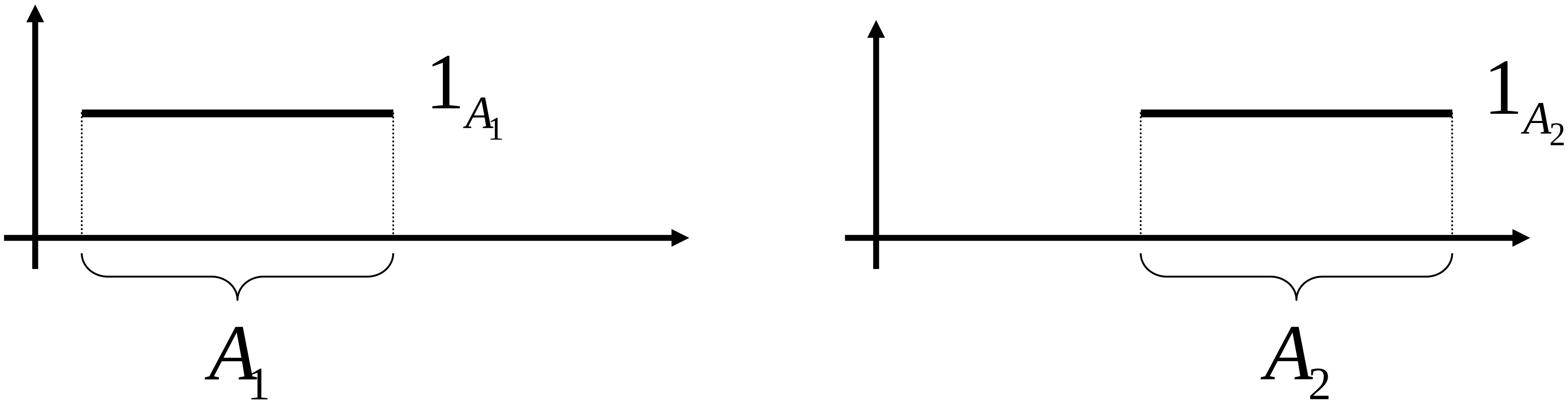
How can we test measurability of a *f?*

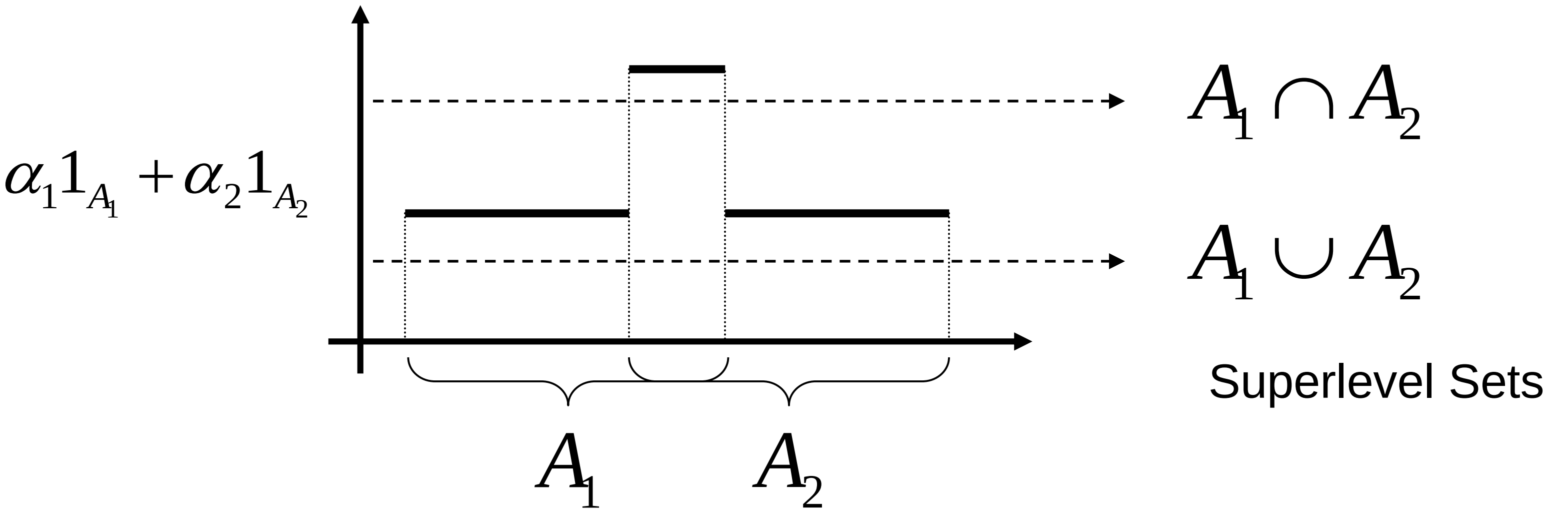
The theorem that helps: $f^{-1}(\sigma(S))=\sigma(f^{-1}(S))$

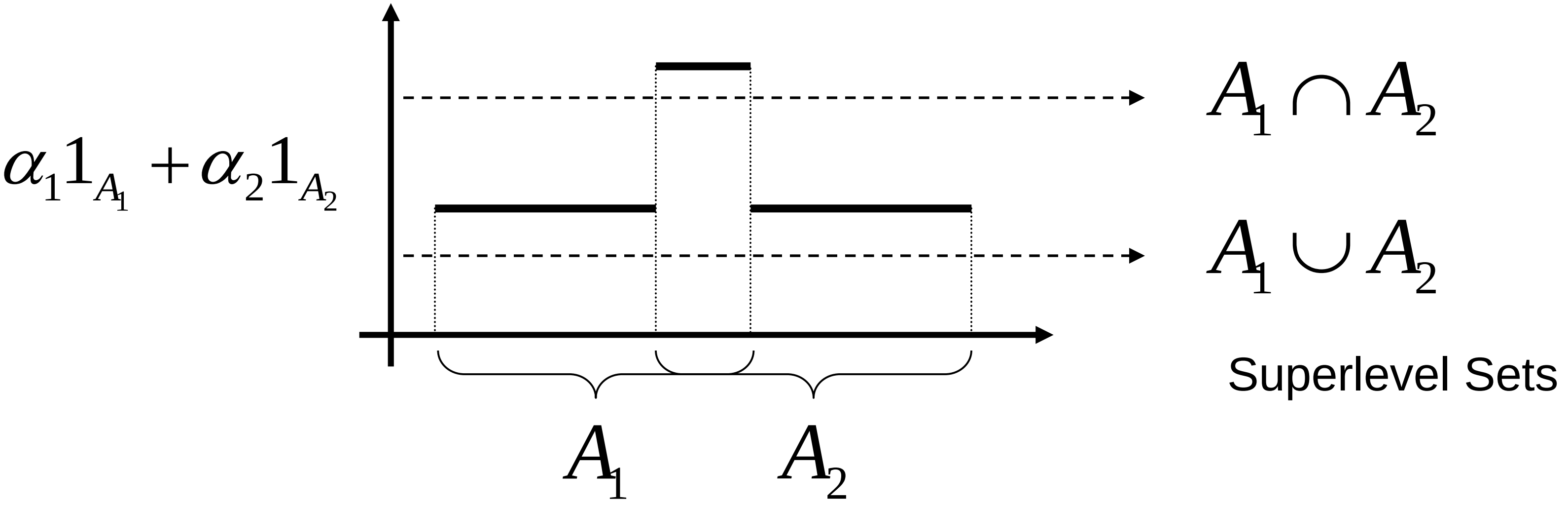
$$\text{If } f^{-1}(S) \subset F_1$$

That implies f is F_1 -F_2 *measurable.*

Algebras are what we need for super/sub-level sets of a vector space of indicator functions.







Simple Functions

Simple functions are finite linear combinations of of indicator functions.

$$\sum_{i=1}^{n} \alpha_i 1_{A_i}$$

σ–algebras let us take limits of these. That is more interesting!!

Measurable Functions

Given a Measurable Space (Ω,*F*)

We say that a function: $y = f(\omega)$ is measurable with respect to *F* if:

$$f : \Omega \rightarrow \Re$$

$$f^{-1}(\{y \geq \alpha\}) \in F \quad \forall \alpha \in \Re$$

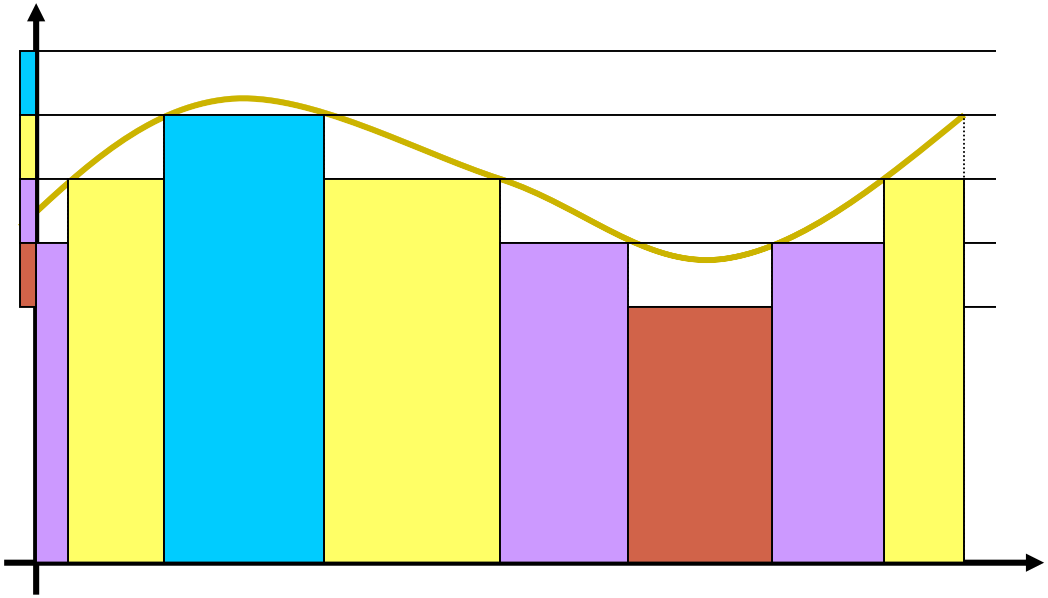
This simply says that a function is measurable with respect to the σ-algebra if all its superlevel sets are in the σ-algebra.

Hence, a measurable function is one that when we slice it like Lebesgue, we can measure the “width” of the part of the function above the slice. Simple...right?

Another equivalent way to think of measurable functions is as the pointwise limit of simple functions.

In fact, if a measurable function is non-negative, we can say it is the increasing pointwise limit of simple functions.

This is simply going back to Lebesgue’s picture...



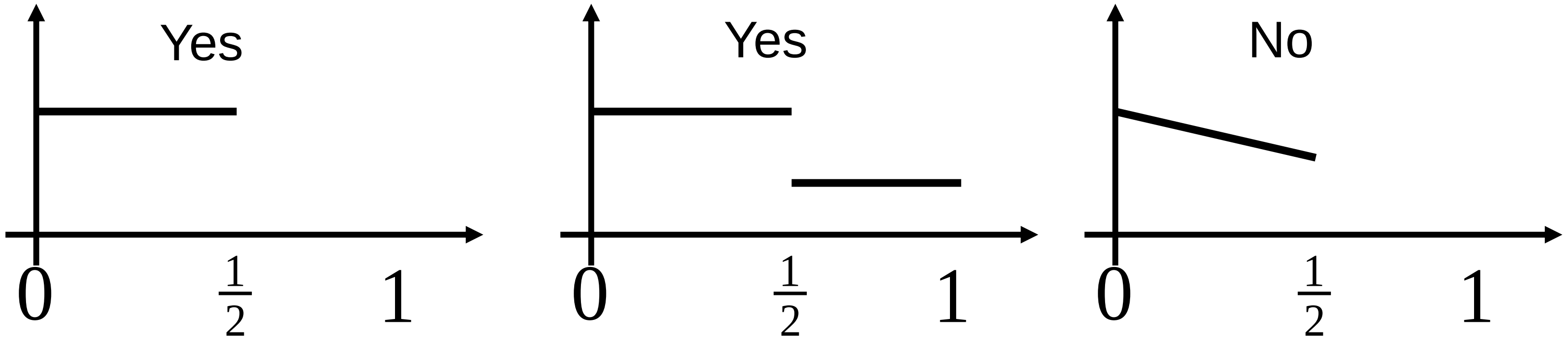
Intuition behind measurable functions.

They are "constant on the sets in the sigma algebra."

$$\Omega = [0,1)$$

$$F = \{\phi, [0,1), [0,\tfrac{1}{2}), [\tfrac{1}{2},1)\}$$

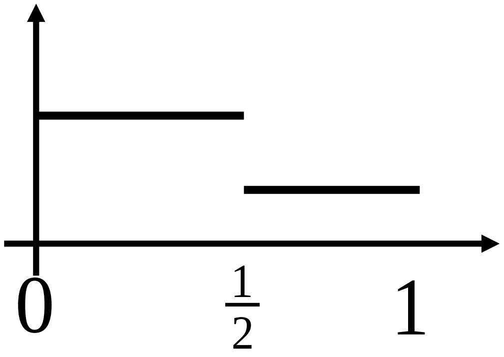
What do measurable functions look like?



Intuitively speaking, Measurable functions are constant on the sets in the σ-alg.

More accurately, they are limits of functions that are constant on the sets in the σ-algebra. The σ in σ-algebra gives us this.

“Information” and s-Algebras.



Since measurable functions are “constant” on the σ-algebra, if I am trying to determine information from a measurable function, the σ-algebra determines the information that I can obtain.

I measure $y = f(\omega)$ What is the most information that I determine about ω?

The best possible I can do is to say that $\omega \in A$ with $A \in F$

σ-algebras determine the amount of information possible in a function.

Measures and Integration

A measurable space (Ω, F) defines the sets can be measured.

Now we actually have to measure them...

What are the properties that "size" should satisfy.

If you think about it long enough, there are really only two...

Definition of a measure:

Given a Measurable Space (Ω, F),

A measure is a function $\mu : F \rightarrow \Re^{+}$

satisfying two properties:

1) $\mu(\phi) = 0$

2) If $A_1, A_2, A_3, \ldots$ are disjoint

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

(2) is known as "countable additivity". However, it can be interpreted as "linearity and left continuity for sets".

Different Measures

- Infinite different measures are available on a sample space

- **Counting measure:** Let μ (A)= n if A contains n number of elements, ∞ otherwise.

- **Discrete measure:** Let $\boldsymbol{\Omega=\{x_1, x_2, \text{-- -} , x_n, \text{-,-} \}}$ **.**

$$\sum_{i=1}^{\infty} p_{x_i} = 1, p_{x_i} \geq 0, \forall i$$

$$\mu(A) = \sum_{x_i \in A} p_{x_i}$$

Different Measures

Lebesgue measure) : There is a unique measure m on (R ,B) that satisfies

m ([*a* , *b*]) = *b* - *a*

for every finite interval [*a* , *b*] , . This is called the *Lebesgue measure* . If we restrict m to measurable space ([0 , 1] , B) , then m is a probability measure

Lebesgue –Stieltjes measure) : Let F be a non-decreasing and right-continuous function from from R to R.There is a unique measure m on (R ,B) that satisfies

m ([*a* , *b*]) = *F(b)* – *F(a)* … … … … …

for every finite interval [*a* , *b*] . This is called the *Lebesgue-Stieltjes measure* . If we take bounded F, then m is a probability measure

Measure-Probability Measure-Cumulative Distribution Function

Bounded measure

Probability measure

Cumulative Distribution Function

Point function Vs Set Function

$$F(x)=Pr\{X<=x\}$$

Point Function

Set Function

What is the relation between them??

Cumulative Distribution Function Vs Probability Function

F(x)||f(x), p(x)

Directly Related to measure

No easy connection with measure

What is the relation between them??

Continuity of Functions

f: $\langle S_1, d_1\rangle \Longrightarrow \langle S_2, d_2\rangle$

is continuous at s in S_1.

$$\Updownarrow$$

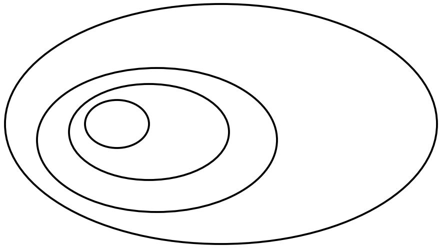
For all $s_n \longrightarrow s \quad \Longrightarrow \quad f(s_n) \longrightarrow f(s)$

it is true in a general metric space but not in general topological space.

Fundamental properties of measures (or “size”):

Left Continuity: This is a trivial consequence of the definition!

Let $A_i \uparrow A$ then $\lim \mu(A_i) = \mu(A)$



Right Continuity: Depends on boundedness!

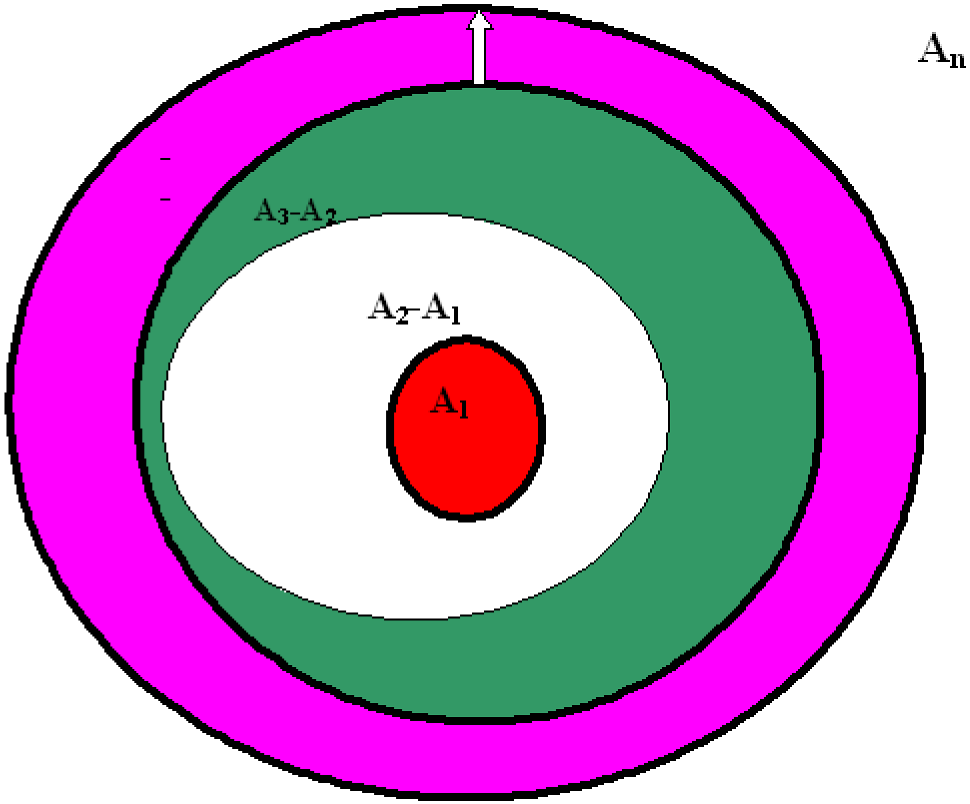
Let $A_i \downarrow A$ and for some i $\mu(A_i) < \infty$

then $\lim \mu(A_i) = \mu(A)$

(Here is why we need boundedness. Consider $A_i = [i, \infty)$ then $A = \phi$)

Continuity –above- property of Measure

$A=A_1 U(A_2-A_1)U------(A_n-A_{n-1})------$



$\mu(A)= \mu(A_1)+ \mu(A_2)- \mu(A_1)+------+ \mu(A_{n\,0}- \mu(A_{n-1})------$

Let $A_i \uparrow A$ then $\lim \mu(A_i) = \mu(A)$

Induced Measures by Measurable Functions

$f\colon\ \Omega_1 \xrightarrow{\hspace{8cm}} \Omega_2$

Ω_1 and Ω_2 are endowed with σ algebras F_1 and F_2 respectively.

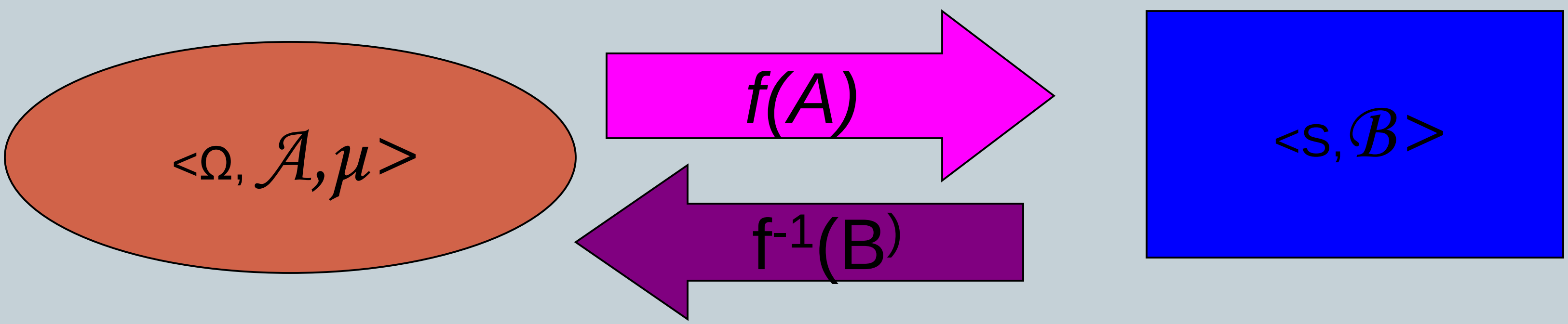
Let us take any B in F_2 and take inverse of B, $f^{-1}\{B\}$. Since B, subset of Ω_2 , $f^{-1}\{B\}$ is a subset Ω_1..

If $f^{-1}\{B\}$ belong to F_1 for all B in F_2 , f is called measurable w.r. t. F_1 and F_2

Induces a Measure on Its Measurable Codomain

- But How?

The domain should be a measure space with a measure. Let μ be the measure on the domain



$m(B)=\mu(f^{-1}(B))$ for all B in $\mathcal{B}$. The definition is possible since $f^{-1}\{B\}$ belong to $\mathcal{A}$. M is written as $\mu \circ f^{-1}$.

General Way to Prove a Theorem for Integration

Prove it for nonnegative simple function

Prove it for nonnegative f^+, f^-

Prove it for $f=f^+-f^-$

Now we can define the integral.

⇨ Simple Functions: $\varphi = \sum_{i=1}^{n} \alpha_i 1_{A_i}$

$$\int \varphi d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

⇨ Positive Functions:

Since positive measurable functions can be written as the increasing pointwise limit of simple functions, we define the integral as

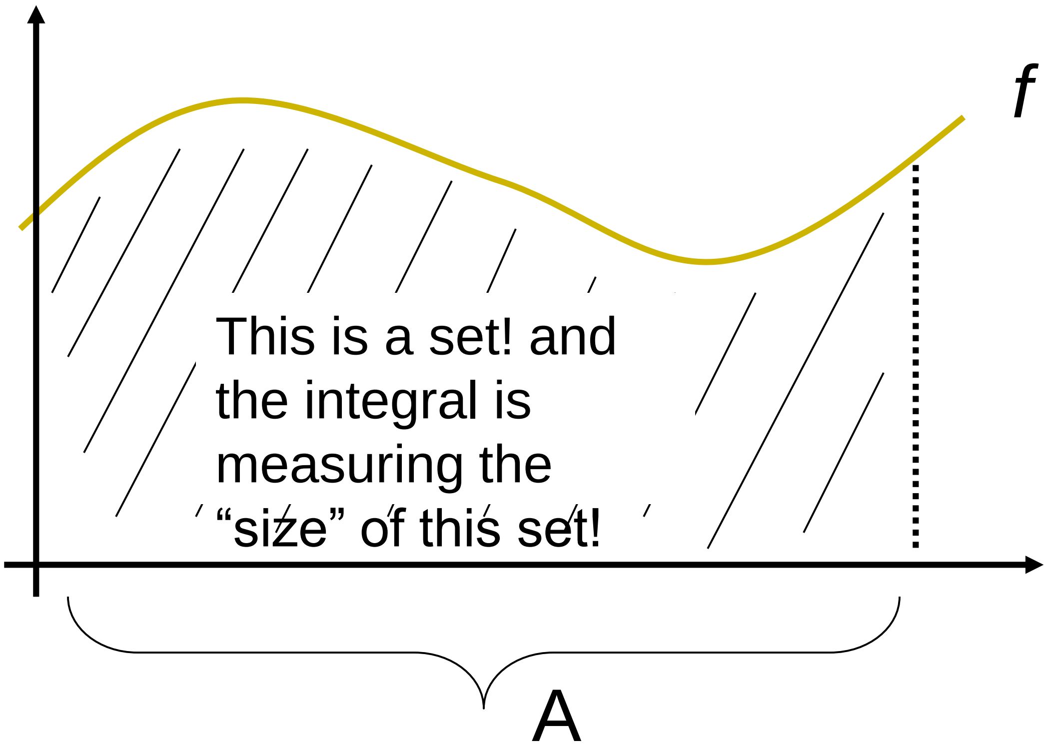
$$\int f d\mu = \sup_{\varphi \leq f} \int \varphi d\mu$$

⇨ For a general measurable function write $f = f^{+} - f^{-}$

$$\int f d\mu = \int f^{+} d\mu - \int f^{-} d\mu$$

How should I think about $\int_A f\,d\mu$

This picture says everything!!!!



Integrals are like measures! They measure the size of a set. We just describe that set by a function.

Therefore, integrals should satisfy the properties of measures.

This leads us to another important principle...

Measures and integrals are different descriptions of of the same concept. Namely, “size”.
Therefore, they should satisfy the same properties!!

Lebesgue defined the integral so that this would be true!

Measures are:	Integrals are:
Left Continuous $A_i \uparrow A \;\Rightarrow\; \lim \mu(A_i) = \mu(A)$	Left Continuous $f_i \uparrow f \;\Rightarrow\; \lim \int f_i d\mu = \int f d\mu$ (Monotone Convergence Thm.)
Bdd Right Cont. $A_i \downarrow A$ and $\mu(A_i) < \infty$ $\Rightarrow\; \lim \mu(A_i) = \mu(A)$	Bdd Right Cont. $f_i \downarrow f$ and $\int f_i d\mu < \infty$ $\Rightarrow\; \lim \int f_i d\mu = \int f d\mu$ (Bounded Convergence Thm.)
etc...	(Fatou's Lemma, etc...) etc...

The L^p Spaces

A function is in $L^p(\Omega,F,\mu)$ if $\int | f |^p \, d\mu < \infty$

The L^p spaces are Banach Spaces with norm:

$$\| f \|_p = \left(\int | f |^p \, d\mu \right)^{1/p} \qquad 1 \le p \le \infty$$

L^2 is a Hilbert Space with inner product:

$$< f , g >= \int fg d\mu$$

Examples of Normed spaces

L_p spaces

- $L_p=\{f:<X, \mathcal{A}, \mu> \rightarrow C \mid (\int |f|^p d\mu)^{1/p} < \infty, 1 <= p < \infty\}$ with $||f|| = (\int |f|^p d\mu)^{1/p}$ is a Banach space if we consider f=g a.e. are equal. For counting measure the condition is not needed,

- C(X) is dense in it if $X=R^k$, $\mathcal{A}=\mathcal{B}(R^k)$ and μ=Lebesgue Stieltjes measure.

- 0<p<1 it is a Fréchet soace

- For p=∞, if define $||f||_\infty$=ess sup|f| <∞ and ess sup g=inf$\{c \in \bar{R} : \mu\{\omega : g(\omega) > c\} = 0\}$, under this norm the space is Banach

Examples of Normed spaces

L_p spaces (Cont.)

- The space is very very large.
- The space is not only important for Kernel methods but also for the development of Fucntional Analysis as well as Foundation of Theory of Statistics
- When X is finite, μ, the counting measure and p=2, we have our well-known, R^k

Checking of Existence of Mean of Cauchy Distribution:

Let X be a standard Cauchy variate with probability density function,

$$f(x) = \frac{1}{\pi(1+x^2)} \; ; -\infty < x < \infty$$

Since X be a continuous random variable, therefore by definition of mean,

$$E(X) = \int xf(x)dx; \quad \text{Provided that the integration exists.}$$

Here we use two way of integral to determine the above integration, they are;

(i) Lebesgue Integral

(ii)Riemann Integral

Checking of Existence of Mean of Cauchy Distribution:

Lebesgue Integral Way

$$\int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x}{1+x^2}dx$$
$$= \frac{1}{\pi}\left[\int_{-\infty}^{\infty} x^+\frac{1}{1+x^2}dx - \int_{-\infty}^{\infty} x^-\frac{1}{1+x^2}dx\right]$$
$$= \frac{1}{\pi}\left[\int_{0}^{\infty} x^+\frac{1}{1+x^2}dx - \int_{-\infty}^{0} x^-\frac{1}{1+x^2}dx\right]$$

Now,

$$\int_{0}^{\infty}\frac{x}{1+x^2}dx > \int_{1}^{\infty}\frac{x}{1+x^2}dx$$
$$= \int_{\substack{\Omega=[1,\infty) \\ =\bigcup[n,n+1)}}\frac{x}{1+x^2}dx$$
$$= \frac{1}{2}\sum_{n=1}^{\infty}\Big[\log(z)\Big]_{1+x^2}^{1+(n+1)^2}$$
$$= \frac{1}{2}\sum_{n=1}^{\infty}\log\frac{1+(n+1)^2}{1+n^2}$$

put

$$1+x^2 = z$$
$$2xdx = dz$$

x	n+1	$1+(n+1)^2$
z	n	$1+n^2$

Checking of Existence of Mean of Cauchy Distribution:

Since $$\log\frac{1+(n+1)^2}{1+n^2} > \log\frac{1+(n+1)^2}{1+2n^2}$$

And $$\lim_{n\to\infty}\log\frac{1+(n+1)^2}{1+n^2} = \frac{1}{2}$$

$$\Rightarrow \int_0^\infty \frac{x}{1+x^2}dx$$ does not exist

$$\Rightarrow E(X)$$ does not exist

Therefore mean of the Cauchy distribution does not exist

Checking of Existence of Mean of Cauchy Distribution:

Riemann Integral Way

Method (i):

We know that, $\lim_{a\to\infty}\int_{-a}^{a}\frac{x}{1+x^2}dx=0$; Since $\frac{x}{1+x^2}$ is an odd function

$$\text{But} \lim_{a\to\infty}\int_{-2a}^{a}\frac{x}{1+x^2}dx$$

$$=\lim_{a\to\infty}\frac{1}{2}\Big[\log(z)\Big]_{1+4a^2}^{1+a^2}$$

$$=\lim_{a\to\infty}\frac{1}{2}\log\frac{1+a^2}{1+4a^2}$$

$$=\frac{1}{2}$$

$\Rightarrow E(X)$ does not exist.

Checking of Existence of Mean of Cauchy Distribution:

Riemann Integral Way

Method (ii):

$$\frac{\frac{x}{1+x^2}}{\frac{1}{x}} = \frac{x^2}{1+x^2} = \lim_{x\to\infty} \frac{x^2}{1+x^2} = 1 \neq 0$$

$\therefore \int_0^\infty \frac{x}{1+x^2}\,dx$ does not exist by quotient test

$\Rightarrow E(X)$ does not exist

Therefore, mean of the Cauchy distribution does not exist.

Measure-Probability Measure-Cumulative Distribution Function

Bounded measure $\Rightarrow$ Probability measure

$\Updownarrow$

Cumulative Distribution Function

Broad Categories of Probability Measures

On Euclidean sample space

Probability Measures

Discrete P(A)=1, #(A)=finite or

$\aleph$

Continuous

P{x}=0 for all x

Absolutely Continous w.r.t. Lebesgue measure

Non A.C.

Expectation

A random variable is a measurable function.

$$X(\omega)$$

The expectation of a random variable is its integral:

$$E(X) = \int X dP$$

A density function is the Radon-Nikodym derivative wrt Lebesgue measure:

$$f_X = \frac{dP}{dx} \Longrightarrow E(X) = \int X dP = \int_{-\infty}^{\infty} x f_X(x) dx$$

Counting measure p=dP/dμ

$=\sum x_i p_i$

Expectation

Change of variables. Let f be measurable from (Ω, F , v) to (Λ,ζ) and g be Borel on (Λ,ζ). Then

$$\int_{\Omega} g \circ f dv = \int_{\Lambda} g d(v \circ f^{-1}) \qquad (1)$$

i.e. , if … r, and the two are the same.

Note that integration domains are indicated on both sides of 1. This result extends the change of variable formula for Reimann integrals, i.e. ,

$$\int g(y)dy = \int g(f(x))f'(x)dx, .y = f(x)$$

Expectation

In statistics we will talk about expectations with respect to different measures.

$$P \longrightarrow E^P(X) = \int X dP$$

$$Q \longrightarrow E^Q(X) = \int X dQ$$

And write expectations in terms of the different measures:

$$E^P(X) = \int X dP = \int X \frac{dP}{dQ} dQ = \int X\varphi dQ = E^Q(X\varphi)$$

where $\frac{dP}{dQ} = \varphi$ or $dP = \varphi dQ$

Thank you