Feb, 19-22, 2015 Kerala School of Mathematics Kozhikode

Measure Theory Refresher Course

B. V. Rao Chennai Mathematical Institute

I. To start with

Imagine Joseph unloading a huge bag of coins (quarter. half, one, two, five, ten rupee coins; if you are familiar you can think of 1,2,3,5,10 paise coins too!) in our class room. He wants us to find out the total value. How do we go about it? We make small heaps (easy) and then each one calculates the total of his/her heap (a little difficult because there are different denominations of coins in your heap). We add all our subtotals. Done. Here is another way. We first separate the coins depending on their value (takes time but no calculation involved) and each of us totals one heap (easy, each heap has coins of one denomination only). Add our subtotals.

Consider the unit interval $[0, 1]$ and a bounded non-negative function f defined on this interval. We want to find area below the graph (bounded by the x-axis). Here is Riemann's prescription. Take an integer $n > 1$. Divide the domain of the function f, namely $[0, 1]$, into small parts $[\frac{k}{n}, \frac{k+1}{n})$ for $k = 0, 1 \cdots, n-1$. For the last interval take right end point too. Denote these parts by $A_0, A_1, \cdots A_{n-1}$. Required area is just sum of the areas over each A_k. So how do you calculate area over A_k? Take any point x_k in A_k and consider the rectangle with base A_k and height $f(x_k)$. *If the function is continuous*, it does not make much difference as to what point is chosen because as n becomes large each of the sets A_k become small and the values of the function on each *fixed* A_k are close to one another. In other words, the quantity $[f(x_k) \times \text{length}(A_k)]$ approximates the area above A_k (under the curve). We add all these areas, name it Riemann sum and take limit as n becomes large. This procedure works if f is continuous function. For example, if $f(x)$ equals zero for rational numbers x and equals one for irrational numbers x, then this procedure fails.

Here is Lebesgue's prescription. Again f is as above defined on $[0, 1]$ and bounded. Since f is non-negative and bounded, let us assume values of f lie in, say $[0, 59]$. Consider the partition of this interval $[0, 59]$ into small parts, $[\frac{k}{n}, \frac{k+1}{n})$ for $k = 0, 1, \cdots, 59n - 1$. For the last interval take right end point too. The set of all points x such that $f(x)$ lies in this interval be denoted by A_k for $0 \leq k \leq 59n - 1$. Thus the interval $[0, 1)$ is again divided into disjoint parts A_k. As above, the total area is the sum of areas above each

A_k. So how do you calculate area above A_k? Again as earlier, take a point $x_k \in A_k$ (if non-empty) and consider the rectangle with base A_k and height $f(x_k)$. Observe that now (without any fuss about continuity of f) the values of f on each fixed A_k are very close to one another — after all, values of f on A_k are between k/n and $(k+1)/n$. Thus $[f(x_k) \times \text{length}(A_k)]$ is a good approximation to the area above A_k. We add all these areas and take limit as n becomes large.

What then is the problem with this second prescription? Well, the sets A_k need no longer be intervals and we have to explain what we mean by length of A_k. This is precisely the problem that Lebesgue settled in an ingenious way. Developing the theory involves several issues: identify a good class of sets for which you can define length; identify a good class of functions f for which the sets A_k defined above do possess length; define integral and explain its properties; finally explain what is new that is achieved by all this.

There were several other developments by 1900 that necessitated the above theory. Emile Borel was curious about the following. Suppose we take a number x, $0 \leq x \leq 1$. consider its decimal expansion. Do each of the digits $\{0, 1, 2, \cdots, 9\}$ occur with the same frequency or is there any injustice to some digits? Clearly you can manufacture plenty of numbers in whose decimal expansion a particular digit does not occur. Borel made the question precise and showed the following. Let A be the set of numbers in whose decimal expansion each digit occurs with frequency $1/10$. He showed that length of A is one. This necessitated understanding length of sets more complicated than intervals. But a clear concept of length for more general sets remained unresolved.

Prior to Borel, Fourier made a profound discovery. You know that the functions $\sin nx$ and $\cos nx$ are periodic with period 2π whatever be $n = 0, 1, 2, \cdots$. He stated that every periodic function is just a superposition of these, that is, linear combination (possibly infinite linear combination) of these functions. This necessitated understanding several issues: the nature of sets where such combinations converge; if two different combinations converge to the same number for a large set of points x then whether the combinations are actually same and so on. all these questions remained unresolved.

Let us recall what the architect Lebesgue says in the introduction to his book 'Lecons sur L' intégration'.

One might ask if there is sufficient interest to occupy oneself with such *complications, and if it is not better to restrict oneself to the study of functions that necessitate only simple definitions. .... As we shall see, in this course, we would then have to renounce the possibility of resolving many problems posed long ago, and which have simple statements. It is to solve these problems, and not for the love of complications, that I have introduced in this book a definition of the integral more general than that of Riemann.*

Without going into the finer details, here are our beginnings. Riemann integral was the main character till about 1900. Motivated by his studies on the patterns of digits in decimal expansion, Emile Borel, in 1898, extended the concept of length for more general sets than intervals and their finite unions. Henri Lebesgue, around 1902, extended length to all Borel sets and defined the Lebesgue integral, thus laying a firm foundation for an epoch making general theory of integration. Soon after in 1905, Vitali constructed his non-measurable set showing certain limitations. Radon realized that the theory can be pushed to more general situations than length, but still living on the Real line. Maurice Frechet around 1915 realized that what you need is a nice class of sets and a nice set function on that class of sets to carry out integration. In 1918, Caratheodory developed his theory of extending measure from a field to σ-field, thus laying a firm foundation for constructing measures. The theorems of F. Riesz (representation of linear functionals), Haar (translation invariant measures on locally compact groups) and developments in functional Analysis/Harmonic Analysis made the theory an essential part of life. S. Saks, Von Neumann and others gave finishing touches and brought the theory to the masses (I mean, popularized through lectures). Around 1923 A N Kolmogorov made measure theory as the foundation for a rigorous development of Probability Theory.

II. Length and measure

Let us consider R, the set of real numbers. We start with a simple collection of sets for which we all agree what is length. The collection consists of the following: R, $\emptyset$, intervals of the form $(a, b]$ where $-\infty < a \leq b < \infty$, intervals $(-\infty, b]$ and intervals of the form (a, ∞). We could have cut short this long sentence if only we used nice convention. We are just considering all intervals of the form $(a, b]$ where $-\infty \leq a \leq b \leq \infty$; the convention being $(a, \infty] = (a, \infty)$. Think about it.

Denote this collection of sets by $\mathcal{S}$. We all agree that length of interval $(a, b]$ is $b - a$; which could be infinity and this happens only when the interval is unbounded. Of course, you might say that we could have considered all kinds of intervals, (a, b) and $[a, b]$ and $[a, b)$ also. Even for these intervals we agree length is $b - a$. Yes, but considering all types of intervals will make our life unmanageable. So we consider only one kind. But at the end you will see that for such intervals also we have assigned length and it agrees with what you have felt.

Theorem 1: Countable additivity on intervals

(A) The collection $\mathcal{S}$ satisfies the following three conditions:

(i) R and $\emptyset$ are in $\mathcal{S}$.

(ii) $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$.

(iii) $A \in \mathcal{S}$ implies A^c is a finite disjoint union of sets in $\mathcal{S}$.

(B) $\lambda((a, b]) = b - a$ is a countably additive set function on the collection $\mathcal{S}$. This means, if $A = \overset{\infty}{\underset{1}{\cup}} A_i$ is a disjoint union where all the sets are in $\mathcal{S}$, then $\lambda(A) = \sum_1^\infty \lambda(A_i)$.

Proof of theorem 1:

Part (A): (i) is part of definition of $\mathcal{S}$. (ii) and (iii) follow from

$$(a, b] \cap (c, d] = (a \vee c, b \wedge d]$$

$$(a, b]^c = (-\infty, a] \quad \cup \quad (b, \infty].$$

Remember convention regarding infinities.

Part (B):

This will be done in a series of three steps.

1^o. If we have a finite number of disjoint intervals $(a_n, b_n]$ for $1 \le n \le k$, all contained in $(a, b]$ then

$$b - a \ge \sum (b_n - a_n).$$

Assume that the intervals are non-empty, thus $a_n < b_n$ for each n. Since we have finitely many intervals, rearrange them if necessary and assume $a_1 < a_2 < a_3 < \cdots < a_k$. Keep in mind the telescopic sum

$$b - a = [b - b_k] + [b_k - a_k] + [a_k - b_{k-1}] + \cdots$$

and ignore the unnecessary terms – all terms being positive – you get the desired inequality. If you are feeling uncomfortable handling infinities, do it in two cases: the interval $(a, b]$ is bounded; it is unbounded. In the first case you have no infinities. In the second case matters are trivial because both sides are visibly infinity.

2^o. If we have intervals (a_n, b_n) for $1 \leq n \leq k$, whose union includes $[a, b]$ then

$$b - a \leq \sum_1^k [b_n - a_n].$$

If one of the intervals (a_n, b_n) is unbounded right side is infinity and no argument is needed.

So assume (a_n, b_n) are all bounded. Of course, $[a, b]$ is also bounded then. The result is clearly true for $k = 1$. Suppose it is true for some k. If you have $k + 1$ intervals take the interval that includes the point b – without loss of generality assume it to be the $(k+1)$-st interval. Observe that the remaining k intervals cover $[a, a_{k+1}]$. Use induction hypothesis and add one more term and simplify. (Incidentally, what happens if $a_{k+1} < a$?).

3^o. Let $I_n = (a_n, b_n]$ be disjoint with union $I = (a, b]$. First assume $(a, b]$ is bounded. Of course then, all the intervals $(a_n, b_n]$ are bounded too. Need to show that

$$b - a = \sum [b_n - a_n].$$

By observation 1^0, the inequality

$$b - a \geq \sum_1^k [b_n - a_n]$$

holds for each finite sum on the right side and hence holds for the total sum. That is,

$$b - a \geq \sum [b_n - a_n].$$

We need to show

$$b - a \leq \sum [b_n - a_n].$$

To get this fix $\epsilon > 0$. Let $\epsilon_n = \epsilon/2^n$. Since the open intervals $(a_n, b_n + \epsilon_n)$ cover the compact interval $[a + \epsilon_1, b]$ we see finitely many of them cover and (2^o) shows sum of lengths of those finitely many intervals is at least $b - a - \epsilon_1$. Thus

$$\sum [b_n - a_n] + \sum \epsilon_n \geq b - a - \epsilon_1$$

That is

$$\sum [b_n - a_n] \geq b - a - 2\epsilon$$

and ϵ being arbitrary, countable additivity follows (what if $a + \epsilon_1 > b$ in this argument?).

Finally, suppose that the interval $(a, b]$ is unbounded, say it is $(-\infty, b]$. Need to show that $\sum[b_n - a_n] = \infty$. This is immediate because for any $-\infty < k < b$, the intervals $(a_n \vee k, b_n]$ are disjoint with union $(k, b]$. (Just consider only those intervals for which $b_n > k]$. So from earlier para

$$\sum[b_n - a_n] \geq \sum[b_n - (a_n \vee k)] = b - k$$

This being true for any $k < b$ you get the desired result. Similar argument works when $a = \infty$.

Let us take a modest step. Let us denote by $\mathcal{F}$ the collection of sets which are finite disjoint union of sets in $\mathcal{S}$. We shall extend concept of length for sets in this larger class in the *obvious* way and show that we still maintain countable additivity.

Theorem 2: countable additivity on field

(A) The collection $\mathcal{F}$ satisfies the following three conditions:

(i) R and $\emptyset$ are in $\mathcal{F}$.

(ii) $A, B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$.

(iii) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$.

(B) If $A \in \mathcal{F}$ and $A = \cup A_i$ finite disjoint union of sets A_i in $\mathcal{S}$ then define $\lambda(A) = \sum \lambda(A_i)$. then λ is well defined on $\mathcal{F}$; gives the same value as above if the set A is already in $\mathcal{S}$ and it is countably additive. This means, if $A = \overset{\infty}{\underset{1}{\cup}} A_i$ is a disjoint union, where all the sets A and $A_i; i \geq 1$ are in $\mathcal{F}$, then $\lambda(A) = \sum_{1}^{\infty} \lambda(A_i)$.

Proof of theorem 2:

Part (A): (i) is already true for $\mathcal{S}$.

If $A = \cup A_i$ finite disjoint union of $\mathcal{S}-$sets and $B = \cup B_j$ finite disjoint union of $\mathcal{S}-$sets then

$$A \cap B = \underset{i,j}{\cup}\{A_i \cap B_j\}$$

finite disjoint union of $\mathcal{S}-$ sets; proving (ii).

If $A = \cup A_i$ finite disjoint union of $\mathcal{S}-$sets, then

$$A^c = \cap A_i^c.$$

Since A_i^c is a finite disjoint union of $\mathcal{S}-$sets, (ii) above completes the proof of (iii).

Part (B):

Define for $A \in \mathcal{F}$ which is a disjoint union of finitely many sets $A_n \in \mathcal{S}$, $\lambda(A) = \sum \lambda(A_n)$. This is a good definition because if the same A is a disjoint union of finitely many sets B_m, $1 \leq m \leq l$ from $\mathcal{S}$ then

$$\sum_n \lambda(A_n) = \sum_n \sum_m \lambda(A_n \cap B_m) = \sum_m \sum_n \lambda(A_n \cap B_m) = \sum_m \lambda(B_m)$$

Here the first equality is from the facts that for each n, A_n is the union of disjoint sets $\{A_n \cap B_m : 1 \leq m \leq l\}$, these are all in $\mathcal{S}$, and λ is additive on this class; second equality is just interchange of the order of sums; and the third equality is from the fact that for each m, B_m is the union (now over n) of the disjoint sets $A_n \cap B_m$.

This λ is actually countably additive on $\mathcal{F}$. Indeed let A_n be disjoint sets from $\mathcal{F}$ whose union, say, A is again in $\mathcal{F}$. Each of these sets in turn are disjoint union of finitely many sets from $\mathcal{S}$, say

$$A_n = \bigcup_j B_{n,j}, \quad n = 1, 2, 3 \cdots; \quad A = \bigcup_i B_i.$$

Of course for different n, you may need different number of sets – that is the range of j in the union may depend on n. Let it be so. Now, using additivity of λ from earlier theorem, and the fact that intersection of two sets in $\mathcal{S}$ is again in $\mathcal{S}$, we get

$$\begin{array}{rcll}
\lambda(A) & = & \sum_i \lambda(B_i) & \text{by definition of } \lambda(A) \\
 & = & \sum_i \sum_n \sum_j \lambda(B_i \cap B_{n,j}) & B_i \text{ is disjoint union over } n, j \text{ of } B_i \cap B_{n,j} \\
 & = & \sum_n \sum_i \sum_j \lambda(B_i \cap B_{n,j}) & \text{interchange of order of sum} \\
 & = & \sum_n \lambda(A_n) & A_n \text{ is disjoint union over } i,\ j \text{ of } B_i \cap B_{n,j}
\end{array}$$

This completes the proof.

We shall now extend the concept of length to a very very large class of subsets of R. This class will not include all subsets of R, in fact, it excludes many many sets. But what ever sets you can think of, are all in our collection!

But before proceeding further let us see the impact of theorem 2. Is it really important that we are dealing with real line? Is it really important that we are considering length?

Suppose Ω is a non-empty set and $\mathcal{S}$ be a collection of subsets of Ω. Say that the collection is a semi-field if it satisfies the three conditions of theorem 1. That is,

(i) Ω and $\emptyset$ are in $\mathcal{S}$.
(ii) $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$.
(iii) $A \in \mathcal{S}$ implies A^c is a finite disjoint union of sets in $\mathcal{S}$.

Let us say that a collection $\mathcal{F}$ is a field of subsets of Ω if it satisfies the three conditions of theorem 2. That is

(i) Ω and $\emptyset$ are in $\mathcal{F}$.
(ii) $A, B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$.
(iii) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$.

Let us say that a function λ defined for sets in $\mathcal{S}$ is a measure if (i) $\lambda(A) \geq 0$ for every set in $\mathcal{S}$; (ii) $\lambda(\emptyset) = 0$ and (iii) λ is countably additive, that is if a set $A \in \mathcal{S}$ is disjoint union $\cup A_i$ of a sequence of sets in $\mathcal{S}$ then $\lambda(A) = \sum \lambda(A_i)$.

Similarly, let us say that a function λ defined for sets in $\mathcal{F}$ is a measure if (i) $\lambda(A) \geq 0$ for every set in $\mathcal{F}$; (ii) $\lambda(\emptyset) = 0$ and (iii) λ is countably additive, that is if a set $A \in \mathcal{F}$ is disjoint union $\cup A_i$ of a sequence of sets in $\mathcal{A}$ then $\lambda(A) = \sum \lambda(A_i)$.

In the proof of theorem 2, we have not used the fact that we have real line. If you look at the proof you will see what we proved is the following.

Theorem 2^*: Extension from semi-field to field

Let λ be a measure on a semifield $\mathcal{S}$ of subsets of Ω.

(A) Let $\mathcal{F}$ be the collection of finite disjoint union of sets in $\mathcal{S}$. Then $\mathcal{F}$ is a field.

(B) Define $\lambda(A) = \sum \lambda(A_i)$ when $A \in \mathcal{F}$ is a finite disjoint union of sets $A_1, \cdots, A_k \in \mathcal{S}$. Then λ is well defined and is a measure on $\mathcal{F}$.

Of course you will wonder if this observation is worth anything at all, because theorem 1 tells length is countably additive on $\mathcal{S}$. Do we have any other examples? Yes, let us understand impact of theorem 1.

Say that a function $F : R \to R$ is a Stieltjes function if (i) it is non-decreasing, that is, $x < y$ implies $F(x) \leq F(y)$ and (ii) it is right continuous, that is, $x_n \downarrow x$ implies $F(x_n) \to F(x)$. Here is what Theorem 1 achieves.

Theorem 1^*: Measures other than length

Let F be a Stieltje's function. Define λ on $\mathcal{S}$ of theorem 1 by $\lambda((a,b]) = F(b) - F(a)$. Then λ is a measure on $\mathcal{S}$.

In fact the same proof as that of Theorem 1 works with obvious modifications. Only in the last part (3^o) we replaced $(a_n, b_n]$ by $(a_n, b_n + \epsilon_n)$. Instead of this replace it by (a_n, b'_n) where $b'_n > b_n$ and $F(b'_n) < F(b_n) + \epsilon_n$ which is possible by the right continuity of F.

Thus for example you could take $F(x) = (x \vee 0) \wedge 1$. Or $F(x) = x^2 \wedge 0$. Draw their graphs.

Let Ω be a non-empty set. Let $\mathcal{A}$ be a collection of subsets of Ω. We say that $\mathcal{A}$ is a $\sigma-$field if the following three conditions hold.
(i) Ω and $\emptyset$ are in $\mathcal{A}$.
(ii) $A_n; n \geq 1$ are all in $\mathcal{A}$ then $\cup A_n \in \mathcal{A}$.
(iii) $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$.

Thus the difference between $\sigma-$field and field is the following. In a field finite unions of sets are there, but an infinite union may not be there. In a $\sigma-$field a countable union of sets, whether finite or infinite union, is there.

Theorem 3: getting $\sigma-$fields
Given any collection $\mathcal{C}$ of subsets of Ω, there is a smallest $\sigma-$field of subsets of Ω which includes all sets in $\mathcal{C}$.

This is denoted by $\sigma(\mathcal{C})$, sigma-field generated by $\mathcal{C}$. When $\Omega = R$ and $\mathcal{C}$ is the collection of sets $\mathcal{S}$ then $\sigma(\mathcal{S})$ is called the Borel sigma field on R and is denoted by $\mathcal{B}$. Sets in $\mathcal{B}$ are called Borel sets.

Proof of theorem 3: Make a list of all $\sigma-$fields which include the given class $\mathcal{C}$. There is at least one such $\sigma-$field, namely, all subsets of Ω. Let $\mathcal{A}$ consist exactly those sets which belong to all the $\sigma-$fields in the above list. (In other words we are considering intersection of all sigma fields which include $\mathcal{C}$).

Theorem 4: Length on Borel $\sigma-$field
There is a unique function λ on $\mathcal{B}$, the Borel sigma field of R, which agrees with length on $\mathcal{S}$ and is countably additive. This means, if $A = \bigcup_1^\infty A_i$ where all the sets (A_i) are in $\mathcal{B}$, then $\lambda(A) = \sum_1^\infty \lambda(A_i)$.

Moreover such a function λ is unique. It is called Lebesgue measure.

This is achieved by proving the following.

Theorem 4^***: Caratheodory Extension Theorem**

Let $\mathcal{F}$ be a field of subsets of a non-empty set Ω and let μ be a measure on $\mathcal{F}$. Assume that there is a sequence of sets Ω_n such that $\mu(\Omega_n) < \infty$ for each n and $\Omega = \cup\Omega_n$.

Then there is a unique extension of μ as a measure to $\sigma(\mathcal{F})$. That is, there is a unique measure μ on $\sigma(\mathcal{F})$ which agrees with the given μ for sets in $\mathcal{F}$.

Obviously this theorem 4^* implies theorem 4 as a special case.

Proof of Theorem 4^* **:** This is one of the few lengthy; but natural and beautiful; proofs.

Uniqueness:

Let us first show that there can not be two extensions.

This is done after observing two simple facts.

(1). Let μ be any measure on $(\Omega, \mathcal{A})$. Then the following is true:

$A_n \uparrow A$, $A_n \in \mathcal{A}$ for all n then $\mu(A_n) \uparrow \mu(A)$.

$A_n \downarrow A$, $A_n \in \mathcal{A}$ for all n, and $\mu(A_1) < \infty$ then $\mu(A_n) \downarrow \mu(A)$.

The first statement is just a reflection of the following fact. If $B_1 = A_1$ and $B_n = A_n - A_{n-1}$ for $n > 1$, then these are disjoint sets in $\mathcal{A}$ whose union is A so that

$$\mu(A) = \sum \mu(B_k) = \lim_{n\to\infty} \sum_1^n \mu(B_k) = \lim_{n\to\infty} \mu(A_n).$$

The second statement is a reflection of the following fact. For $n \geq 1$ put $B_n = A_n - A_{n+1}$. Then these are disjoint sets in $\mathcal{A}$ and $\cup\{B_k : k \geq n\} = A_n - A$. In particular $\sum_1^\infty \mu(B_k) = \mu(A_1 - A) < \infty$. So tail sums $\sum_n^\infty \mu(B_k)$ converge to zero, but this tail sum is $\mu(A_n - A)$. Thus $\mu(A_n) \to \mu(A)$. (Where did we use $\mu(A_1) < \infty$?)

(2) There is no loss to assume that the sets Ω_n in the theorem are disjoint. This is because, if they are not, you can replace by $\Omega_1' = \Omega_1$ and $\Omega_n' = \Omega_n - \bigcup_1^{n-1} \Omega_k$ for $n > 1$.

Returning to uniqueness, let μ and ν be two extensions of μ given on $\mathcal{F}$. We show for each $A \in \mathcal{A}$ and $n \geq 1$; $\mu(A \cap \Omega_n) = \nu(A \cap \Omega_n)$. By adding over

n; see (2) above; we get $\mu(A) = \nu(A)$ for each $A \in \mathcal{A}$ as desired.

So fix n. Let $\mathcal{M}$ be the collection of sets $A \in \mathcal{A}$ for which $\mu(A \cap \Omega_n) = \nu(A \cap \Omega_n)$. This collection includes the field $\mathcal{F}$ and is; by (1) above; a monotone class. Hence by the monotone class theorem this class must equal $\mathcal{A}$ completing the proof of uniqueness.

Existence, special case $\mu(\Omega) = 1$:
If $\mu(\Omega) = 1$ then the measure μ is called a probability.
We start with fine tuning of observation (1) above.
$A_n \uparrow A$, $A_n \in \mathcal{F}$ for all n and $A \in \mathcal{F}$ then $\mu(A_n) \uparrow \mu(A)$. (♠)
In observation (1) above we have taken sets from the $\sigma-$field. Now we are taking sets from the field. Proof is exactly as earlier using countable additivity. Note that now we are assuming that A is in the field.

Plan of the proof is the following. We first make a modest extension of the given measure to a large collection of sets. Next we try to understand the maximum value we can assign to an arbitrary set. In the final step we collect those sets for which the maximum value we can assign equals the minimum value we must assign. Thus there is nothing left to our choice! This does.

Step 1: Let

$$\mathcal{F}_\sigma = \{A : \text{ there is a sequence of sets } A_n \in \mathcal{F}; A_n \uparrow A\}.$$

Of course, since $\mathcal{F}$ is a field, this class is nothing but countable unions of sets in $\mathcal{F}$. Guided by (♠) above, for A in this class put $\mu(A) = \lim_n \mu(A_n)$. Since the sequence A_n is increasing, this limit exists. This is a good definition because if you take another sequence B_n from the field increasing to A, then for every m, $(B_n \cap A_m) \uparrow A_m$ and all these sets are in $\mathcal{F}$ so that by (♠), $\mu(B_n \cap A_m) \uparrow \mu(A_m)$ and hence, for every m,

$$\lim_n \mu(B_n) \geq \lim_n \mu(B_n \cap A_m) = \mu(A_m).$$

True for all m, implying $\lim_n \mu(B_n) \geq \lim_m \mu(A_m)$. Similarly one proves $\lim_m \mu(A_m) \geq \lim_n \mu(B_n)$.

Note that if $A \in \mathcal{F}$ then we can take each A_n to be A itself to see that the value $\mu(A)$ defined in the earlier para is same as the starting value. Thus what we have defined is an extension of the given μ. Thus using the same symbol μ for this extension does not lead to any confusion. This extension

μ takes values in $[0,1]$. It satisfies three properties that we need later. Here they are.

(1 a) It is monotone — that is, if $A \subset B$ then $\mu(A) \leq \mu(B)$.

Indeed if $A_n \uparrow A$ and $B_n \uparrow B$ then $A_n \cap B_n$ is in $\mathcal{F}$ for each n and $(A_n \cap B_n) \uparrow A$ so that $\mu(A) = \lim \mu(A_n \cap B_n) \leq \lim \mu(B_n) = \mu(B)$.

(1 b) If $A_n \in \mathcal{F}_\sigma$ and $A_n \uparrow A$, then $A \in \mathcal{F}_\sigma$ and $\mu(A_n) \uparrow \mu(A)$.

Indeed let $A_{m,n} \uparrow_m A_n$ for each n where all these sets $A_{m,n}$ are in $\mathcal{F}$. Put $B_n = \bigcup_{i,j \leq n} A_{i,j}$. observe that this is finite union of sets in $\mathcal{F}$ and is hence in $\mathcal{F}$ again. Note also that $B_n \uparrow A$. This first of all shows that $A \in \mathcal{F}_\sigma$ and $\mu(A) = \lim \mu(B_n)$. Using the monotonicity of the sequence of sets A_n, observe that $B_n \subset A_n$ so that $\mu(B_n) \leq \mu(A_n)$. Thus

$$\mu(A) = \lim \mu(B_n) \leq \lim \mu(A_n) \leq \mu(A)$$

the last inequality is from (a), because $A_n \subset A$ for all n. This shows that $\mu(A) = \lim \mu(A_n)$.

(1 c) If A and B are in $\mathcal{F}_\sigma$ then $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$. Indeed if $A_n \uparrow A$ and $B_n \uparrow B$ then $(A_n \cap B_n) \uparrow (A \cap B)$ and $(A_n \cup B_n) \uparrow (A \cup B)$ showing that both $A \cap B$ and $A \cup B$ are in $\mathcal{F}_\sigma$. Since the sets A_n, B_n, are all in $\mathcal{F}$ we also have

$$\mu(A_n) + \mu(B_n) = \mu(A_n \cup B_n) + \mu(A_n \cap B_n)$$

Taking limits and using the definition of our extension we get the stated equality.

Step 2:

For every subset $B \subset \Omega$ let us put

$$\mu^*(B) = \inf\{\mu(A) : B \subset A \in \mathcal{F}_\sigma\}.$$

If at all we want to extend our measure this is the maximum value we can give for the set B. Clearly this μ^* also takes values in the interval $[0,1]$.

If B is already in $\mathcal{F}_\sigma$ then note that B itself is one candidate for the set A in the definition and any other A would have a possibly larger value by (1 a), showing that $\mu^*(B) = \mu(B)$. Thus μ^* extends μ from $\mathcal{F}_\sigma$. This μ^* defined on all subsets of Ω has three properties that we need later.

(2 a) It is monotone — that is, if $B_1 \subset B_2$, then $\mu^*(B_1) \leq \mu^*(B_2)$. This is immediate because any set that includes B_2 includes B_1 as well.

(2 b) For any two sets B_1 and B_2;

$$\mu^*(B_1) + \mu^*(B_2) \geq \mu^*(B_1 \cup B_2) + \mu^*(B_1 \cap B_2)$$

To see this, fix $\epsilon > 0$, pick for $i = 1, 2$; $A_i \supset B_i$, A_i in $\mathcal{F}_\sigma$ such that $\mu(A_i) \leq \mu^*(B_i) + \epsilon$. Then

$$\mu^*(B_1) + \mu^*(B_2) + 2\epsilon \geq \mu(A_1) + \mu(A_2) = \mu(A_1 \cup A_2) + \mu(A_1 \cap A_2)$$

The equality above is a consequence of (1c). But since $A_1 \cup A_2 \supset B_1 \cup B_2$ and $A_1 \cup A_2 \in \mathcal{F}_\sigma$ we deduce that $\mu(A_1 \cup A_2) \geq \mu^*(B_1 \cup B_2)$ and similarly $\mu(A_1 \cap A_2) \geq \mu^*(B_1 \cap B_2)$. Using this in the display above, we get

$$\mu^*(B_1) + \mu^*(B_2) + 2\epsilon \geq \mu^*(B_1 \cup B_2) + \mu^*(B_1 \cap B_2)$$

Since ϵ is arbitrary, the statement is verified.

(2 c) If $B_n \uparrow B$ then $\mu^*(B_n) \uparrow \mu^*(B)$.

Of course by (2a) we know that $\mu^*(B_n)$ is increasing and is at most $\mu^*(B)$. Thus $\lim_n \mu^*(B_n) \leq \mu^*(B)$. Towards proving the reverse inequality, fix $\epsilon > 0$. For each $n \geq 1$, fix $A_n \supset B_n$, $A_n \in \mathcal{F}_\sigma$ and $\mu(A_n) \leq \mu^*(B_n) + \epsilon/2^n$. Set $C_n = \cup_{m \leq n} A_m$ so that C_n are increasing, $C_n \in \mathcal{F}_\sigma$, $C_n \supset B_n$. We claim that for each n,

$$\mu(C_n) \leq \mu^*(B_n) + \sum_{m \leq n} \frac{\epsilon}{2^m}$$

This is clear, by choice of A_1, for $n = 1$. Going by induction, suppose we did upto n. Using (1c),

$$\mu(C_{n+1}) = \mu(C_n \cup A_{n+1}) = \mu(C_n) + \mu(A_{n+1}) - \mu(C_n \cap A_{n+1})$$

Now induction hypothesis and choice of A_{n+1} gives

$$\mu(C_{n+1}) \leq \mu^*(B_n) + \sum_{m \leq n} \frac{\epsilon}{2^m} + \mu^*(B_{n+1}) + \frac{\epsilon}{2^{n+1}} - \mu(C_n \cap A_{n+1})$$

Since $C_n \supset A_n \supset B_n$ and $A_{n+1} \supset B_{n+1} \supset B_n$ we see that $C_n \cap A_{n+1}$ is a set in $\mathcal{F}_\sigma$ which includes B_n so that $\mu(C_n \cap A_{n+1}) \geq \mu^*(B_n)$. Using this in the display above we get

$$\mu(C_{n+1}) \leq \mu^*(B_{n+1}) + \sum_{m \leq n+1} \frac{\epsilon}{2^m}$$

to complete the inductive step.

If $C = \cup_n C_n$, then $C_n \uparrow C$. By (1b), $C \in \mathcal{F}_\sigma$,

$$\mu(C) = \lim_n \mu(C_n) \leq \lim_n \mu^*(B_n) + \epsilon$$

from what we observed above. Since $C \supset B$, this implies

$$\mu^*(B) \leq \mu(C) \leq \lim_n \mu^*(B_n) + \epsilon$$

and ϵ being arbitrary we get $\mu^*(B) \leq \lim_n \mu^*(B_n)$ completing the proof.

Step 3;

It is natural now to get a lower bound for the value $\mu(B)$ that we can assign to B and then collect those sets for which both the upper and lower bounds coincide – because for such sets there is nothing we should decide! But a moment's reflection shows that the lower bound for the value of $\mu(B)$ is nothing but $1 - \mu^*(B^c)$. Thus the condition that both upper and lower bounds coincide is the same as saying $\mu^*(B) + \mu^*(B^c) = 1$. Let $\mathcal{B}$ be the collection of all sets B satisfying this condition, that is, $\mu^*(B) + \mu^*(B^c) = 1$. We now understand this class $\mathcal{B}$ and properties of μ^* on this class.

(3 a) We first show that $\mathcal{B}$ is a field and μ^* on this class is additive.

By the very definition of the class $\mathcal{B}$ it is clear that if B is in this class then so is B^c and hence $\mathcal{B}$ is closed under complements. To show $\mathcal{B}$ is closed under finite unions, take A and B in $\mathcal{B}$. So,

$$\mu^*(A) + \mu^*(A^c) = 1; \quad \mu^*(B) + \mu^*(B^c) = 1 \tag{1}$$

Apply (2 b) to A and B,

$$\mu^*(A) + \mu^*(B) \geq \mu^*(A \cup B) + \mu^*(A \cap B) \tag{2}$$

Apply (2 b) to A^c and B^c,

$$\mu^*(A^c) + \mu^*(B^c) \geq \mu^*((A \cup B)^c) + \mu^*((A \cap B)^c) \tag{3}$$

Add (2) and (3) and use (1)

$$2 \geq \mu^*(A \cup B) + \mu^*(A \cap B) + \mu^*((A \cup B)^c) + \mu^*((A \cap B)^c) \tag{4}$$

Apply (2b) to $A \cup B$ and $(A \cup B)^c$

$$\mu^*(A \cup B) + \mu^*((A \cup B)^c) \geq 1 \tag{5}$$

Apply (2b) to $A \cap B$ and $(A \cap B)^c$

$$\mu^*(A \cap B) + \mu^*((A \cap B)^c) \geq 1 \tag{6}$$

Strict inequality in either (5) or (6) will $--$ after adding (5) and (6) $--$ contradict (4). Hence equality must hold in both (5) and (6), which means that both $A \cup B$ and $A \cap B$ are in $\mathcal{B}$.

Note that equality in (5) and (6) deduced above implies that equality holds in (4) as well. Since strict inequality in either (2) or (3) will again lead to strict inequality in (4), we conclude that equality holds in (2) and (3) as well. In particular, we have

$$\mu^*(A) + \mu^*(B) = \mu^*(A \cup B) + \mu^*(A \cap B)$$

All this amounts to saying that $\mathcal{B}$ is a field and μ^* is finitely additive on this class $--$ as claimed.

(3b) We now show that $\mathcal{B}$ is a σ-field and μ^* is countably additive on this.

We first do the following: take $B_n \uparrow B$ where each $B_n \in \mathcal{B}$ and show that $B \in \mathcal{B}$ and $\mu^*(B_n) \uparrow \mu^*(B)$. But this last statement is already (2c), so we only need to show $B \in \mathcal{B}$, that is $\mu^*(B) + \mu^*(B^c) = 1$. Of course, by (2b), $\mu^*(B) + \mu^*(B^c) \geq 1$. So we only need to show $\mu^*(B) + \mu^*(B^c) \leq 1$. Now, since each $B_n \in \mathcal{B}$ we have $\mu^*(B_n) + \mu^*(B_n^c) = 1$. But $B_n \subset B$ so that $B_n^c \supset B^c$. Consequently, $\mu^*(B^c) \leq \mu^*(B_n^c)$ Thus $\mu^*(B_n) + \mu^*(B^c) \leq 1$. Take limit over n and use (2c) to get the desired inequality.

Since we showed that μ^* is finitely additive on $\mathcal{B}$ the observation of earlier para shows countable additivity as follows. Take disjoint sets $A_n \in \mathcal{B}$ and put $B_n = \bigcup_{k=1}^{n} A_k$. Then $B_n \uparrow \cup A_k$ so that,

$$\mu^*(\cup A_k) = \lim \mu^*(B_n) = \lim \sum_1^n \mu^*(A_k) = \sum_1^\infty \mu^*(A_k).$$

(3c) Since μ is a measure on $\mathcal{F}$; for $B \in \mathcal{F}$ we have $\mu(B) + \mu(B^c) = 1$. Since μ^* extends μ, the same holds for μ^* as well. In other words, $\mathcal{F} \subset \mathcal{B}$. Now (b) above implies that $\sigma(\mathcal{F}) \subset \mathcal{B}$ and μ^* is a measure on this which extends the given μ on $\mathcal{F}$. To complete the proof of the theorem, we only need to look at μ^* on $\sigma(\mathcal{F})$.

Existence, special case $\mu(\Omega) < \infty$:

If $\mu(\Omega) < \infty$ then the measure is called a finite measure. The previous case immediately yields this because if $\mu(\Omega) = c$ then $\nu(A) = \mu(A)/c$ for $A \in \mathcal{F}$ defines a measure and if we consider its extension ν' to $\mathcal{A}$, then clearly $\mu' = c\nu'$ will do. Of course, if $c = 0$ there is nothing to be done.

Existence:

When there is a sequence of sets $\Omega_n \in \mathcal{F}$ such that $\mu(\Omega_n) < \infty$ for each n and $\Omega = \cup\Omega_n$, then the measure is called a $\sigma-$finite measure on $\mathcal{F}$.

In this case, as observed earlier, you can take the sets Ω_n to be disjoint. Then $\mu_n(A) = \mu(A \cap \Omega_n)$ for $A \in \mathcal{F}$ then this is a finite measure and if μ'_n is its extension to $\mathcal{A}$ then $\sum \mu'_n$ extends μ.

This completes the proof.

Can we define length for all subsets of R?

Theorem 5: Impossibilty of defining for all subsets

(Under a set theoretic hypothesis) It is impossible to extend length defined on $\mathcal{S}$ to all subsets of R, so that it is countably additive.

This is not difficult, but shall not spend time on this.

III. Integral

Suppose that $(\Omega, \mathcal{A}, \mu)$ is a measure space. Suppose that s is a simple non-negative measurable function say $s = \sum a_i I_{A_i}$ where for each i, $a_i \geq 0$ and the sets A_i are disjoint, $\cup A_i = \Omega$. We define

$$\int s d\mu = \sum a_i \mu(A_i).$$

We take $0 \times \infty = 0$.

If f is a non-negative measurable function then we put

$$\int f d\mu = \sup\{\int s d\mu : s \text{ simple measurable}, 0 \leq s \leq f\}.$$

If f is any measurable function, we say that f is integrable iff

$$\int f^+ d\mu < \infty; \quad \text{and} \quad \int f^- d\mu < \infty$$

and in such a case we put

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

We denote the collection of integrable functions by $L^1(\Omega, \mathcal{A}, \mu)$, in short, L^1 when other things are understood.

Why is integral defined this way

The first question that should be answered is the reasons to adapt this definition. Suppose we have $s = I_A$, then our definition gives $\int s = \mu(A)$. This stands to reason because s takes two values one and zero and hence the area under s consists of two rectangles: one of unit height with base length $\mu(A)$ and the other of zero height with base length $\mu(A^c)$. Thus these two rectangles have total area $\mu(A)$.

Since integral is linear for non-negative simple functions it should be as defined above.

Obviously if we have two non-negative functions $f \leq g$ then area under g is at least as large as area under f. Thus in particular for every simple non-negative $s \leq f$ we must have $\int s \leq \int f$. Thus $\int f$ should not fall below the value we have prescribed. The best way at this stage is to see if we can get an upper bound for the possible value of $\int f$ and then decide what to do. If luckily these two values coincide then there is nothing for us to decide. Though this path can also be pursued carefully, we did not do. There is no *a priori* reason why $\int f$ should exceed this lower bound (if so by how much? etc). Moreover you have learnt that there are enough simple nonnegative measurable functions below f to 'cover' all its area. Thus it appears sensible to declare this lower bound itself as $\int f$. This is what we did.

Of course, a general measurable function f can be expressed as $f = f^+ - f^-$ so that linearity justifies the definition we gave. Just note the following. Integral is defined for every non-negative measurable function, at the worst it could be infinity. However, even to define integral for general f we first define concept of integrability and then assign integral for such functions. This is necessitated because we do not have a meaning for $\infty - \infty$.

Of course it is not clear whether the integral so defined has reasonable properties at all. The next theorem says that all properties you expect from integral hold with this definition.

Theorem 6: Properties of integral.

(i) If a simple nonnegative measurable function s is expressed in two different ways as

$$s = \sum a_i I_{A_i}; \quad a_i \geq 0; \quad A_i \cap A_j = \emptyset \text{ for } i \neq j; \quad \cup A_i = \Omega$$

$$s = \sum b_j I_{B_j}; \quad b_j \geq 0; \quad B_i \cap B_j = \emptyset \text{ for } i \neq j \quad \cup B_j = \Omega.$$

Then

$$\sum a_i \mu(A_i) = \sum b_j \mu(B_j).$$

In other words $\int s$ for simple nonnegative measurable functions is well defined. $\int s$ is nonnegative linear in the sense that for simple nonnegative s and t and non-negative reals α and β we have

$$\int (\alpha s + \beta t) = \alpha \int s + \beta \int t.$$

Also, $\int s$ is monotone in the sense that if $s \leq t$ then $\int s \leq \int t$.

(ii) The clause defining integral for non-negative measurable functions extends the definition given for simple non-negative measurable functions earlier. $\int f$ is monotone on the collection of non-negative measurable functions.

(iii) (**Monotone convergence theorem ; MCT**) If $f_n \geq 0$ and $f_n \uparrow f$ then $\int f_n \uparrow \int f$.

(iv) If $f \geq 0$, $g \geq 0$ and $\alpha, \beta \geq 0$ then $\int (\alpha f + \beta g) = \alpha \int f + \beta \int g$.

(v) The final clause defining integral for general (integrable) measurable functions extends the definition for non-negative integrable measurable functions.

Further, $f \in L^1$ iff $|f| \in L^1$. Also, $|\int f| \leq \int |f|$.

L^1 is a vector space and integral is linear on this space and is also monotone.

(vi) (**Fatou's lemma**) $\int \liminf f_n \leq \liminf \int f_n$ for nonnegative measurable functions f_n.

(vii) (**Lebesgue's Dominated convergence theorem ; DCT**) If $f_n \in L^1$ for each n; $f_n \to f$ point wise ; if there is $h \in L^1$ such that $|f_n| \leq h$

point wise for all n, then $f \in L^1$ and $\int |f_n - f| \to 0$ and $\int f_n dP \to \int f$.

(viii) Let f be a non-negative measurable function. $\mu(f \neq 0) = 0$ iff for every set $A \in \mathcal{A}$, $\int f I_A d\mu = 0$. More precisely, if $\mu(f \neq 0) = 0$ then f is integrable, $\int |f| = 0$ and for every $A \in \mathcal{A}$, the function $f I_A$ is integrable and integral zero. Conversely, if f is integrable and $\int f I_A = 0$ for every set $A \in \mathcal{A}$ then $\mu(f \neq 0) = 0$.

As a consequence, for two measurable functions f and g, $\mu(f \neq g) = 0$ iff for every set $A \in \mathcal{A}$, $\int f I_A d\mu = \int g I_A d\mu$.

Proof of Theorem 6 :

(i) Observe that $a_i \mu(A_i \cap B_j) = b_j \mu(A_i \cap B_j)$ for each i, j. This is because if $\mu(A_i \cap B_j) \neq 0$ then take a point ω in it and observe that $a_i = f(\omega) = b_j$.

The first statement follows from additivity of μ

$$\sum_i a_i \mu(A_i) = \sum_i \sum_j a_i \mu(A_i \cap B_j) = \sum_j \sum_i b_j \mu(A_i \cap B_j) = \sum b_j \mu(B_j).$$

Suppose that

$$s = \sum_1^n a_i I_{A_i}, \quad t = \sum_1^m b_j I_{B_j}$$

then clearly

$$s + t = \sum_{i,j} (a_i + b_j) I_{(A_i \cap B_j)}.$$

Now compute integrals

(ii) Just to avoid confusion, temporarily use $\int^*$ for expectation defined for non-negative measurable functions. Let f be simple nonnegative measurable function. Note that f itself is a candidate for s in the definition of $\int^* f$. Thus $\int^* f \geq \int f$. On the other hand if we take any other non-negative simple $s \leq f$ then monotonicity of expectation, observed in (i) above, yields that $\int s \leq \int f$ showing that $\int^* f = \int f$. Monotonicity of expectation is immediate from definition.

(iii) Start observing that if s is a simple nonnegative random variable and Ω_n are sets in the σ-field increasing to Ω then $E(sI_{\Omega_n}) \uparrow E(s)$. This is done by direct computation. If $s = \sum c_i I_{B_i}$ (finite sum), then $sI_{\Omega_n} = \sum c_i I_{B_i \cap \Omega_n}$ and hence

$$\int (sI_{\Omega_n}) = \sum c_i \mu(B_i \cap \Omega_n) \uparrow \sum c_i \mu(B_i) = E(s)$$

For the proof of MCT, let $f_n \uparrow f$ – all being nonnegative. We already know that $\int f_n \leq \int f$ and increases with n , so that $\lim \int f_n \leq \int f$. Towards the other inequality fix any simple nonnegative $s \leq f$. Suffices to show that $\int s \leq \lim \int f_n$. Again fix $0 < \alpha < 1$. Suffices to show that $\int (\alpha s) \leq \lim \int f_n$. Let $\Omega_n = \{\omega : \alpha s(\omega) \leq f_n(\omega)\}$, so that $\Omega_n \uparrow \Omega$. By monotonicity of integral, we have $\int (\alpha s I_{\Omega_n}) \leq \int (f_n I_{\Omega_n}) \leq \int f_n$. Take limits and use the observation made at the beginning to complete the proof.

(iv) Get simple nonnegative $s_n \uparrow f$ and $t_n \uparrow g$. From (i), $\int (s_n + t_n) = \int (s_n) + \int (t_n)$, use MCT for all the three terms noting that $s_n + t_n \uparrow f + g$ to complete the proof.

(v) That the definition extends from non-negative variables is clear since, if $f \geq 0$ then $f^+ = f$ and $f^- = 0$ so that the new definition is same as the old one. First note that f is integrable iff both $\int (f^+)$ and $\int (f^-)$ are finite – equivalently,$\int |f| = \int f^+ + \int f^-$ is finite. thus note that $f \in L^1$ iff $|f| \in L^1$. Further by definition, $|\int f| = |\int f^+ - \int f^-|$ is at most $\int f^+ + \int f^- = \int |f|$.

In particular, if f and g are integrable, then so are $|f|$ and $|g|$ and hence their sum because of additivity of integral on non-negative measurable functions. Since integral is monotone on non-negative measurable functions and since $|f + g| \leq |f| + |g|$ we conclude that $\int (|f + g|)$ is finite and hence $f + g$ is integrable. Finally, since

$$(f+g)^+ - (f+g)^- \quad = f + g = \quad f^+ - f^- + g^+ - g^-$$

we conclude that

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$$

Now take intgral on both sides and since all terms are finite, rearrange them carefully to get $\int (f + g) = \int f + \int g$. Monotonicity is routine.

(vi) If $g_n = \inf\{f_n, f_{n+1}, \cdots\}$ then $g_n \leq f_n$. So $\int g_n \leq \int f_n$ for all n which implies that $\liminf \int g_n \leq \liminf \int f_n$. Note that g_n are increasing to $\liminf f_n$ and so by MCT, the left side is actually $\int \liminf f_n$ to complete the proof.

(vii) The hypotheses imply $|f| \leq h$ and hence f is integrable. So are $|f_n - f|$. The non-negative measurable functions $2h - |f_n - f|$ converge point wise to $2h$ an application of Fatou gives

$$\int 2h d\mu \leq \liminf \int (2h - |f_n - f|) d\mu.$$

Linearity of integral and finiteness of the integrals involved gives after cancellation of the h integral, $\limsup \int |f_n - f| d\mu \leq 0$. Since $\int |f_n - f| d\mu$ are positive numbers we conclude

$$\lim \int |f_n - f| d\mu \leq 0.$$

and since

$$|\int f_n d\mu - \int f d\mu| \leq \int |f_n - f| d\mu$$

we conclude the last sentence as well. This completes the proof.

(viii) Assume that $\mu(f \neq 0) = 0$. In particular $\mu(f^+ \neq 0) = 0$. We show $\int f^+ = 0$. Take any simple non-negative measurable function $s \leq f^+$. Say $s = \sum a_i I_{A_i}$ with $a_i \geq 0$ and A_i disjoint sets. Now $s \leq f^+$ so that if $a_i > 0$ then $\mu(A_i) = 0$. Thus $\int s = 0$. This being true for every simple $0 \leq s \leq f$ we conclude $\int f^+ = 0$.

Similarly $\int f^- = 0$ giving $\int f = 0$. In fact $\int |f| = 0$. For any set A, we have

$$|\int f I_A| \leq \int |f| I_A \leq \int |f| = 0.$$

Conversely assume for every set A, $\int f I_A$ =0. In particular if $B = \{f > 0\}$ and $C = \{f < 0\}$ we see taking A to be these sets that $\int |f| = 0$.

For last sentence, use $f - g$.

IV. Some fine tuning

There are some important details that need to be attended to. There are no new ideas – only extensions of the ideas discussed so far.

infinities

We need to discuss measurable functions that may take values $\pm\infty$ also. Not that we invite them, but they may crop up in our calculations. We realized it while dealing with $\limsup$ and $\liminf$ of measurable functions.

Extended real line means the set $\{-\infty\} \cup R \cup \{+\infty\}$, with the understanding that $-\infty < x < +\infty$ for all real numbers x. Usually $+\infty$ is also denoted as ∞. A function taking values in this set is said to be an extended real valued function. Such a function f is said to be an extended real measurable function if for every real number x, $\{\omega : f(\omega) \leq x\} \in \mathcal{A}$ – as in the case of real measurable functions. We have the common sense arithmetic on this extended real line. for every real x,

$$\pm\infty + x = \pm\infty;$$

$$\infty + \infty = \infty; \quad -\infty - \infty = -\infty.$$

We agree NOT to talk about $\infty - \infty$.

With this arithmetic we can define sums of two extended measurable functions provided at no point ω, one of them assumes value ∞ and the other assumes the value $-\infty$. Then the sum will also be an extended real measurable function. lim sup and lim inf are all (extended real) measurable functions. Integration can also be defined. Of course one still takes the same old definition for simple measurable functions – that is, the values $\pm\infty$ are not allowed for simple measurable functions.

complex measurable functions

Many calculations involve complex valued functions. Bringing in complex numbers puts at our disposal all results from complex analysis. When t is a real number e^t can take any positive value and is hence an unbounded function. On the other hand when t is a real number e^{it} takes values only on the unit circle and is hence a bounded function. Such an advantage helps one to use complex numbers.

Note that if f is a complex valued function on a set Ω, then there are two uniquely defined real valued functions f_1 and f_2 such that

$$f(\omega) = f_1(\omega) + i f_2(\omega).$$

If we have a σ-field $\mathcal{A}$ on Ω then we say that f is a measurable function iff both f_1 and f_2 are so. We already know when to say a real valued function is a measurable function. If moreover we have a measure μ on $\mathcal{A}$ then we say that f is integrable if both f_1 and f_2 are so and then we define

$$\int f = \int f_1 + i \int f_2.$$

All the results for expectation remain valid. Of course there is no linear order on complex numbers and hence MCT does not make sense. But the fact $|E(f)| \leq E|f|$ (with a tricky proof) and the DCT remain valid.

almost everywhere

While dealing with measurable functions we should not ignore the existence of measure. For instance $f \leq g$ if for every $\omega \in \Omega$, $f(\omega) \leq g(\omega)$ is a statement that does not take into account the existence of a measure. From a measure theoretic point of view, if we have a property that depends on points $\omega \in \Omega$ then we should agree that the property holds provided the set of points where the property fails is an unimportant set; that is, it has measure zero. This is what we discuss now.

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. In what follows we consider measurable functions.

Say that $f \leq g$ *a.e.* if $\mu[\omega : f(\omega) > g(\omega)] = 0$.

Say that $f = g$ *a.e.* if $\mu[\omega : f(\omega) \neq g(\omega)] = 0$.

Say that $f_n \uparrow$ *a.e.* if $\mu[\omega : f_n(\omega) > f_{n+1}(\omega) \text{ for some } n] = 0$.

Say that $f_n \to f$ *a.e.* if $\mu[\omega : f_n(\omega) \not\to f(\omega)] = 0$.

Here *a.e.* is abbreviation for almost everywhere. Observe that to make sense of these definitions, we must make sure that the sets in braces are indeed events. But once you recognize the problem, it is easy to sort out.

Most of the theorems on integration can be fine-tuned by putting *a.e.*

For example, MCT can be restated as follows : if $f_n \uparrow f$ *a.e.* and $f_n \geq 0$ *a.e.* then $\int f_n \uparrow \int f$.

Proof is easy. If N_1 is the set of points where the sequence $f_n(\omega)$ is not increasing; N_2 is the set of points where some $f_n(\omega) < 0$ then by hypothesis both these sets have measure zero. As a result if we set $A = (N_1 \cup N_2)^c$ and $g_n = f_n I_A$, $g = f I_A$ then

$$\mu(f_n \neq g_n) = 0; \quad \mu(f \neq g) = 0; \quad g_n \geq 0; \quad g_n \uparrow g$$

Now usual MCT tells you $\int g_n \uparrow \int g$ and last part of theorem 6 tells you $\int g_n = \int f_n$ and $\int g = \int f$ giving the desired result.

DCT can be restated as follows : If $f_n \in L^1$ for each n; $f_n \to f$ *a.e.* ; there is $h \in L^1$ such that $|f_n| \leq h$ *a.e.* for all n, then $f \in L^1$ and $\int |f_n - f| \to 0$ and $\int f_n \to \int f$.

Similar proof as above applies, consider appropriate set A so that integrals do not change but usual DCT applies for $g_n = f_n I_A$.

V Examples

(A) Let $\Omega = \{1, 2, 3, \cdots\}$ and $\mathcal{A}$ be the collection of all subsets of Ω and μ be the counting measure, that is, $\mu(A)$ is number of points of A when A is finite; and equals ∞ when A is an infinite set. Then μ is a measure.

Functions defined on Ω are just sequences $a = \{a_n : n \geq 1\}$. a_n is the value of the function a at the point n. Every function is measurable.

If a function a is non-negative then $\int a d\mu = \sum a_n$. This is because the function f^k defined by $f^k(i) = a_i$ for $i = 1, 2, \cdots k$ and $f^k(i) = 0$ for $i > k$ can be regarded as a simple function and by direct evaluation we see $\int f^k d\mu = \sum_1^k a_n$. Also $f^k \uparrow a$. Thus by MCT

$$\int a d\mu = \lim \int f^k d\mu = \lim \sum_1^k a_n = \sum_1^\infty a_n$$

Now it is easy to see that a function a is integrable iff $\sum |a_n|$ converges and then $\int a d\mu = \sum a_n$

Thus summation is also an integration with respect to a measure.

(B) Let us consider $\Omega = R$, $\mathcal{B}$ the Borel sigma field and λ be Lebesgue measure or length measure. consider a function f which is zero outside $[0, 1]$ and is continuous on $[0, 1]$. Then the Riemann integral $\int_0^1 f(x)dx$ is same as the integral $\int f d\lambda$ as developed above (called Lebesgue integral). We can see this as follows. Take the $1/n$ partition of the interval $[0, 1]$, namely, $\{0, 1/n, 2/n, \cdots, n/n = 1\}$, take a point ξ_k from $[(k-1)/n, k/n)$ (include right end point for the last interval) and consider the Riemann sum

$$\sum_1^n f(\xi_k)/n$$

By definition these converge to the Riemann integral. On the other hand the above Riemann sum can also be regarded as integral of the following simple function: s_n whose value in the interval $[(k-1)/n, k/n)$ is $f(\xi_k)$ and zero outside $[0, 1]$. It is easy to see that $s_n \to f$. Also they are dominated by the integrable function $cI_{[0,1]}$ where c is a bound for the continuous function $|f|$ on $[0, 1]$. so by DCT, these integrals converge to $\int f d\lambda$. Thus the two integrals are same.

However for functions not supported on a bounded interval (or for unbounded functions even though supported on a bounded interval) one has to be a little careful. For example, one says the Riemann integral

$$\int_0^\infty \frac{\sin x}{x} dx = \pi/2.$$

However the Lebesgue integral

$$\int_0^\infty \frac{\sin x}{x} d\lambda$$

does not exist. The Riemann integral above is, what is usually called, an improper Riemann integral (nothing improper about it). Its meaning in the Riemann integral context is defined to be

$$\lim_{A\to\infty} \int_0^A \frac{\sin x}{x} dx$$

Interpreted this way it is true in the Lebesgue integral sense too. that is

$$\lim_{A\to\infty} \int \frac{\sin x}{x} I_{[0,A]}(x) d\lambda = \frac{\pi}{2}.$$

(C) You can construct several examples of measures as follows. Take any integrable non-negative function h on R. Put

$$\mu(A) = \int hI_A d\lambda, \qquad A \in \mathcal{B}.$$

Then DCT tells you that this is a measure. Starting with indicator functions you can show that a function f is $\mu-$integrable iff the function fh is Lebesgue integrable and then $\int f d\mu = \int fh d\lambda$.

(D) Here is another way of getting measures. Suppose that we have a set Ω' and a σ-field $\mathcal{A}'$ of its subsets, and a map T defined on Ω with values in Ω' such that for every set $B \in \mathcal{A}'$, we have $T^{-1}(A') \in \mathcal{A}$. Then we can define a measure on $\mathcal{A}'$ by putting $\nu(B) = \mu(T^{-1}B)$. We get a new measure space $(\Omega', \mathcal{A}', \nu)$. This ν is called the induced measure on $\mathcal{A}'$ induced by μ via T.

Starting with indicator functions of sets in $\mathcal{A}'$ we can show the following. A (measurable) function f on Ω' is $\nu-$integrable iff the composed function foT defined on Ω by $foT(\omega) = f(T(\omega))$ is $\mu-$integrable and then

$$\int f d\nu = \int foT d\mu.$$

this is also called change of variable formula. This is useful even when you take both spaces to be reals R with Borel sigma field.

(E) If you combine the above two discussions (C) and (D) you will get the usual Jacobian formula.

(F) You can construct new measures by taking products, just like defining area using lengths.

VI. Radon-Nikodym

Suppose that we have a $\sigma-$finite measure space $(\Omega, \mathcal{A}, \mu)$. That is, $\mathcal{A}$ is a $\sigma-$field of subsets of Ω and μ is a measure on this. And Ω is a union of sets in $\mathcal{A}$ with finite measure. Let us take a non-negative $\mu-$integrable function f. Define

$$\nu(A) = \int f I_A d\mu; \quad A \in \mathcal{A}.$$

Then ν is again a measure. This can be seen using DCT, s mentioned above. This measure has an interesting property, namely, if $\mu(A) = 0$ then $\nu(A) = 0$ as well, simply because then the integrand is zero a.e. This measure ν is a finite measure.

In fact even if f is not integrable, as long as it is non-negative real valued measurable function the above formula defines a measure ν which is $\sigma-$finite and satisfies:

$$A \in \mathcal{A}; \;\; \mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0 \quad (\clubsuit).$$

The converse is also true, that is, if ν is a $\sigma-$finite measure satisfying $(\clubsuit)$ then it arises in this way. In others words there is a non-negative measurable function f such that $\nu(A) = \int f I_A d\mu$ for all $A \in \mathcal{A}$. Since the property $(\clubsuit)$ is important we shall name it and then prove this statement.

Here is a very useful notation. $\int f I_A d\mu$ is also denoted $\int\limits_A f d\mu$.

Suppose that μ and ν are two measures on $(\Omega, \mathcal{A})$. Say that ν is absolutely continuous w.r.t μ, in symbols $\nu \ll \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$. Say that ν is singular w.r.t μ, in symbols $\nu \perp \mu$, if there is a set $A \in \mathcal{A}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$. Thus the two measures are 'concentrated' on disjoint sets. This relation is symmetric. That is, if $\mu \perp \nu$ then $\nu \perp \mu$. Obviously absolute continuity is not symmetric.

Theorem 7 : Radon-Nikodym and Lebesgue Decomposition Theorems

(i) Let μ and ν be two $\sigma-$finite measures on $(\Omega, \mathcal{A})$. Then $\nu \ll \mu$ iff there exists a measurable function h such that for all $A \in \mathcal{A}$, $\nu(A) = \int\limits_A h\, d\mu$.

Further such a h is almost surely unique – that is, if h_1 and h_2 satisfy the condition then $\mu(h_1 \neq h_2) = 0$. Of course then $\nu(h_1 \neq h_2) = 0$ as well.

Such a h is almost surely w.r.t μ (and hence w.r.t. ν as well) nonnegative.

(ii) Let μ and ν be two $\sigma-$finite measures on $(\Omega, \mathcal{A})$. Then ν is sum of two uniquely defined measures ν_1 and ν_2 with $\nu_1 \ll \mu$ and $\nu_2 \perp \mu$.

Part (i) above is called the *Radon-Nikodym Theorem*. The function h is called the *Radon-Nikodym derivative* of ν w.r.t μ — in symbols $h = \frac{d\nu}{d\mu}$, one also writes $d\nu = h\, d\mu$. Starting with indicator functions, simple functions, non-negative functions, one can show the following. A measurable function f is $\nu-$integrable iff the product fh is $\mu-$integrable and then

$$\int f d\nu = \int f h d\nu.$$

Part (ii) of the above theorem is called the *Lebesgue decomposition Theorem*. The measure ν_1 is called the *absolutely continuous part* of ν and ν_2 is called the *singular part* of ν. These are, of course, relative to μ.

Proof of Theorem 7 (Radon-Nikodym Theorem):

One part was already commented on above, namely, starting with h if you define ν then it is a measure and $\nu \ll \mu$. We now assume that $\nu \ll \mu$ and obtain h as stated.

case μ and ν both finite:

The idea is to look at the collection of all non-negative measurable functions g such that $\int\limits_A g d\mu \leq \nu(A)$ holds for all $A \in \mathcal{A}$. There is at least one such, namely, the zero function. The hope is that there is a largest such function and if we use that function equality must hold above for all sets.

To execute this idea, let

$$L = \{X \geq 0 : \forall A \in \mathcal{A}; \int_A X d\mu \leq \nu(A)\}.$$

Let $\alpha = \sup\{\int X d\mu : X \in L\}$. If X and Y are in L then so is $X \vee Y$. This is because if $S = \{X \geq Y\}$ then for any $A \in \mathcal{A}$ we have

$$\int_{A\cap S} X d\mu \leq \nu(A \cap S); \quad \int_{A\cap S^c} Y d\mu \leq \nu(A \cap S^c)$$

and adding the two inequalities we get $\int_A X \vee Y d\mu \leq \nu(A)$. If $X_n \uparrow X$ and each $X_n \in L$ then $X \in L$. So we can get a $Z \in L$ with $\int Z d\mu = \alpha$. Indeed using the definition of α, for each n you can get $Z_n \in L$ such that

$$\int Z_n \geq \alpha - \frac{1}{n}.$$

By taking successive maximums if necessary, and observing that they are still in L, we can assume that $Z_n \uparrow$. Let $Z_n \uparrow Z$. Then $Z \in L$ and by MCT $\int Z = \alpha$. We first claim that if $X \in L$ then $X \leq Z$ a.e μ. If this were not true then just note that $X \vee Z \in L$ and would have integral larger than α, contradicting the definition of α.

Define a finite non-negative measure on $\mathcal{A}$ by

$$\lambda(A) = \nu(A) - \int_A Z d\mu.$$

If only we could show that $\lambda = 0$ then we are done.

There are exactly two possibilities.
(i) either there is an integer $k \geq 1$ and a set $A \in \mathcal{A}$ such that

$$\lambda(A) > \frac{1}{k}\mu(A).$$

(ii) Or for every integer $k \geq 1$ and every set $A \in \mathcal{A}$

$$\lambda(A) \leq \frac{1}{k}\mu(A).$$

The second alternative implies $\lambda(A) \leq \frac{1}{k}\mu(A)$ for all k and A showing that λ is actually the zero measure; completing the proof as noted earlier.

In the first case we arrive at a contradiction to the fact that Z is the 'largest' function in L. Accordingly fix a k and A as above, that is,

$$\lambda(A) > \frac{1}{k}\mu(A).$$

In particular $\lambda(A) > 0$, so $\nu(A) > 0$ and hence $\mu(A) > 0$. How nice if *for every* $B \subset A$, $B \in \mathcal{A}$ we had $\lambda(B) \geq \frac{1}{k}\mu(B)$. Then we could argue that

$Z + \frac{1}{k} I_A$ is in L and larger than Z on the set A of positive μ measure, contradicting an earlier statement about Z. If A does not have this property, starting from A, we exhibit another set with this extra property.

Set $A_0 = A$ and

$$\beta_0 = \inf\{\lambda(B) - \frac{1}{k}\mu(B) : B \subset A \text{ and } B \in \mathcal{A}\}.$$

If $\beta_0 \geq 0$ we are done. So let $\beta_0 < 0$. Take $B_0 \subset A_0$, $B_0 \in \mathcal{A}$ with $\lambda(B_0) - \frac{1}{k}\mu(B_0) < \frac{1}{2}\beta_0$. Set $A_1 = A_0 - B_0$. Note that if $B \subset A_0 - B_0$ then we must have $\lambda(B) - \frac{1}{k}\mu(B) \geq \beta_0/2$. Otherwise take a set B_1 for which opposite inequality holds, namely

$$\lambda(B_1) - \frac{1}{k}\mu(B_1) < \beta_0/2$$

Then $B_0 \cup B_1 \subset A$ and

$$\lambda(B_0 \cup B_1) - \frac{1}{k}\mu(B_0 \cup B_1) < \beta_0$$

contradicting the definition of β_0.

Also observe

$$\lambda(A_1) - \frac{1}{k}\mu(A_1) = \lambda(A_0 - B_0) - \frac{1}{k}\mu(A_0 - B_0)$$

$$= [\lambda(A_0) - \frac{1}{k}\mu(A_0)] - [\lambda(B_0) - \frac{1}{k}\mu(B_0)]$$

$$> \lambda(A_0) - \frac{1}{k}\mu(A_0)$$

Thus we have
(i) $\lambda(A_1) - \frac{1}{k}\mu(A_1) > \lambda(A_0) - \frac{1}{k}\mu(A_0)$ and
(ii) if $B \subset A_1$ then $\lambda(B) - \frac{1}{k}\mu(B) \geq \beta_0/2$.

Do to A_1 what you did to A_0 . That is, set

$$\beta_1 = \inf\{\lambda(B) - \frac{1}{k}\mu(B) : B \subset A_1 \text{ and } B \in \mathcal{A}\}$$

From what was observed above $\beta_1 \geq \beta_0/2$. If $\beta_1 \geq 0$ then we are done as earlier. So assume that $\beta_1 < 0$. Thus

$$\beta_0/2 \quad \leq \quad \beta_1 < 0.$$

Take $B_1 \subset A_1$, $B_1 \in \mathcal{A}$ with $\lambda(B_1) - \frac{1}{k}\mu(B_1) < \frac{1}{2}\beta_1$. Set $A_2 = A_1 - B_1$. Note that if $B \subset A_1 - B_1$ then we must have

$$\lambda(B) - \frac{1}{k}\mu(B) \geq \beta_1/2 \geq \beta_0/2^2.$$

Also observe

$$\lambda(A_2) - \frac{1}{k}\mu(A_2) > \lambda(A_1) - \frac{1}{k}\mu(A_1) > \lambda(A_0) - \frac{1}{k}\mu(A_0)$$

Thus we have
(i) $\lambda(A_2) - \frac{1}{k}\mu(A_2) > \lambda(A_0) - \frac{1}{k}\mu(A_0)$ and
(ii) if $B \subset A_2$ then $\lambda(B) - \frac{1}{k}\mu(B) \geq \beta_0/2^2$.

Continue. Just note that at each stage either the process stops and you have a set you wanted, or, you have $A_n \subset A_{n-1}$ such that
(i) $\lambda(A_n) - \frac{1}{k}\mu(A_n) > \lambda(A_0) - \frac{1}{k}\mu(A_0)$ and
(ii) if $B \subset A_n$ then $\lambda(B) - \frac{1}{k}\mu(B) \geq \beta_0/2^n$.

If you stopped at a finite stage then you got a set A with the property stated in the earlier para to complete proof. If you continue for ever then put $A_\infty = \cap A_n$. Since properties (i) and (ii) hold at each stage and the sets $A_n \downarrow A_\infty$ we see
(i) $\lambda(A_\infty) - \frac{1}{k}\mu(A_\infty) > \lambda(A_0) - \frac{1}{k}\mu(A_0)$¿0 and
(ii) if $B \subset A_\infty$ then $\lambda(B) - \frac{1}{k}\mu(B) \geq 0$.

This set satisfies what you wanted.

Uniqueness of such a Z is easy using last part of the theorem 6. This completes the proof when both μ and ν are finite.

The general case is as follows. Fix disjoint Ω_n with both μ and ν values finite and $\Omega = \cup\Omega_n$. Define $\nu_n(A) = \nu(A \cap \Omega_n)$ and $\mu_n(A) = \nu(A \cap \Omega_n)$ for $A \in \mathcal{A}$. These are finite measures and for each n, we do have $\nu_n \ll \mu_n$ and hence there is h_n such that $d\nu_n = h_n d\mu_n$. That is for every $A \in \mathcal{A}$

$$\mu(A \cap \Omega_n) = \mu_n(A) = \int_A h_n d\mu_n = \int_A h_n I_{\Omega_n} d\mu.$$

It is not difficult to see that

$$h = \sum h_n I_{\Omega_n}$$

will serve our purpose.

Theorem 7 (Lebesgue decomposition):

First again consider the case of finite measures μ and ν.

Note that $\nu << \mu + \nu$; get the Radon-Nikodym derivative Z. Observe that $Z \leq 1$ w.r.t all these measures. Put

$$\Omega_1 = [0 \leq Z < 1]; \qquad \Omega_2 = [Z = 1].$$

$$\nu_1(A) = \nu(A \cap \Omega_1); \qquad \nu_2 = \nu(A \cap \Omega_2)$$

This is the required decomposition.

First note that ν_1 and ν_2 are measures

Next note that, since Ω_1, Ω_2 is a decomposition of Ω so that $\nu = \nu_1 + \nu_2$.

$$\nu(\Omega_2) = \int_{\Omega_2} Z d(\nu + \mu) = \int_{\Omega_2} 1 d(\mu + \nu) = \mu(\Omega_2) + \nu(\Omega_2).$$

showing $\mu(\Omega_2) = 0$. Since $\nu_2(\Omega_1) = 0$ we conclude that $\nu_2 \perp \mu$.

Finally we show that $\nu_1 \ll \mu$. Let $\mu(A) = 0$. Need to show $\nu_1(A) = 0$. But we already know $\nu_1(\Omega_2) = 0$, we need to show that $\nu_1(A \cap \Omega_1) = 0$. But $\nu_1(A \cap \Omega_1)\nu(A \cap \Omega_1)$

$$\nu(A \cap \Omega_1) = \int_{A \cap \Omega_1} Z d(\mu + \nu) = \int_{A \cap \Omega_1} Z d\nu.$$

the last equality being a consequence of the assumption $mu(A) = 0$. Since $Z < 1$ on Ω_1 we see that the above equation can hold only if $\nu(A \cap \Omega_1) = 0$ as required.

Let if possible there be two decompositions:

$$\nu = \nu_1 + \nu_2; \quad \nu_1 \ll \mu; \quad \nu_2 \perp \mu$$

$$\nu = \nu_1^* + \nu_2^*; \quad \nu_1^* \ll \mu; \quad \nu_2^* \perp \mu.$$

Let $d\nu_1 = f d\mu$ and $d\nu_1^* = f^* d\mu$ Let S and S^* be the sets

$$\mu(S) = 0, \ \ \nu_2(S^c) = 0; \qquad \mu(S^*) = 0, \ \ \nu_2^*(S^{*c}) = 0.$$

Observe that $\mu(S \cup S^*) = 0$ so that

$$A \subset S \cup S^* \ \ \Rightarrow \int_A (f - f^*) d\mu = 0.$$

If $A \subset S^c \cap S^{*c}$ then $\nu_2^*(A) - \nu_2(A) = 0$, that is $\nu_1(A) - \nu_1^*(A) = 0$. Thus

$$A \subset (S \cup S^*)^c \ \ \Rightarrow \int_A (f - f^*) d\mu = 0.$$

Thus for all A we see $\int_A f d\mu = \int_A f^* d\mu$ showing $\nu_1 = \nu_1^*$. Of course, then $\nu_2 = \nu_2^*$.

This completes the proof in the case when μ and are finite. The $\sigma-$finite case is now routine.