A gentle introduction to Measure Theory

Gaurav Chandalia
Department of Computer Science and Engineering
SUNY - University at Buffalo, Buffalo, NY
gsc4@buffalo.edu

March 12, 2007

Abstract

This note introduces the basic concepts and definitions of measure theory relevant to probability theory. It is meant to be a simplified tutorial on measure theory.

1 Introduction

The *Method of Exhaustion* is a technique to find out the area of a shape (for example, circle) by approximating (lower bound) it by inscribing it using a polygon. The idea is that as the number of sides of the polygon tends to infinity the area of the polygon will get arbitrarily close to the area of the circle. This method was used by Archimedes to find out the value of π. Thus, by *exhausting* the values of the area of the polygon, the area of the circle can be obtained. In fact, the Riemann integral is an application of the method of exhaustion.

Measure theory is concerned with generalizing the notions of area on arbitrary sets of Euclidean spaces and notions of length of subsets of $\mathbb{R}$. Essentially, it is a common ground for analysis of real functions and set theory. For the notion of area, there are certain properties like non-negativity and additivity that hold true. Similarly, in measure theory there is a stronger assumption of countable additivity. Measure theory involves study of σ-algebras (abstract mathematical structures), measures, measurable functions and integrals. Integration in the context of measure theory involves analogous sums and is based on functions constant on sets of some σ-algebra and not on functions constant on intervals (as is done in the traditional manner).

I will clearly define a space so that there is something to measure in Sections 2.3 and 2.4; then I will introduce the notion of a measure on this space in Section 3 and show how measures are used to define an integral in Section 3.2.

2 Basics

Before looking at measure theory more formally, I will review some of the basic concepts including some ideas from abstract algebra that define the underlying mathematical structures on which measures are defined.

2.1 Limit of a sequence

A sequence $x_1, x_2, \ldots x_n$ is said to converge to the point or has *limit* x if for every $\epsilon > 0$ there is a natural number N_ϵ such that the neighborhood of $O(x, \epsilon)$ (a circle with center x and radius ϵ) has all the points x_n with $n > N_\epsilon$. Intuitively, it means that the elements of the sequence eventually become arbitrarily close to the point x.

2.2 Cauchy Sequence

A sequence $x_1, x_2, \ldots x_n$ is said to be a *Cauchy sequence* if for every $\epsilon > 0$ there exists an N_ϵ such that $d(x_{n'}, x_{n''}) < \epsilon$ for all $n^{'}, n^{''} \geq N_\epsilon$. Intuitively, it means that the elements of the sequences eventually become close to each other. However, Cauchy sequences are not the same as convergent sequences (having a limit) except in certain cases (for example in $\mathbb{R}$). Some of the important properties of Cauchy sequences are:

- Any Cauchy sequence is bounded.
- Any Cauchy sequence in $\mathbb{R}$ is convergent.

2.3 Algebraic Structure

In pure mathematics, an algebraic structure is a collection of one or more sets that are operated upon by one or more **operations**. These operations satisfy certain **axioms**. An algebraic structure consists of all the possible models that are characterized by the axioms. Algebraic structures are categorized by the number of sets and binary / unary operations they require. Following are some of the relevant structures:

- Groups - A group $(S, *)$ is a set S under a binary operation $* : S \times S \rightarrow S$ that satisfies certain axioms like associativity, closure, e.g. Integers under addition which also happens to be an *abelian group*. An abelian groups is a group in which the order of operation does not matter, i.e. $a * b = b * a$.
- Rings - A ring is a set S under two binary operations : addiction and multiplication. A ring under addition is an abelian group. e.g. integers, polynomials. A commutative ring is a ring in which the members satisfy the commutative law.
- Fields - A field is a structure under which the operations of addition, multiplication, division and subtraction are performed. A field is a commutative ring.

2.4 σ-algebra and Borel algebra

σ-algebra $\mathbb{S}$ is a collection of subsets of a set S that is closed under countable set operations i.e. the complement of a member (subset), the union or intersection of countably many members are also members. Formally, an algebra $\mathbb{S}$ of subsets of a set S is a σ-algebra if $\mathbb{S}$ contains the limit of every monotone sequence of its sets. The pair $(S, \mathbb{S})$ is then a *measurable space* and the sets in S are said to be *measurable*. e.g. 2^S.

Borel algebra of a set S is the minimal σ-algebra that contains all the open sets (or closed sets) on the real line. The elements of Borel algebra are called **Borel Sets**.

σ-algebras (in particular, Borel algebras) allow us to concentrate on sets that are (in some way) easy to use. In other words, these mathematical structures hold certain important properties (countable additivity - 2.5) that allow us to define the concept of a measure on seemingly arbitrary sets.

2.5 Countably Additive

Let ϕ be a real function, and S be a σ-algebra. The function ϕ is countably additive or σ-additive when

$$\phi(S) = \sum_{i=1}^{\infty} \phi(S_i) \quad where \;\; S \in \mathbb{S}\,; \;\; S = \bigcup_{i=1}^{\infty} S_i\,; \;\; S_i \cap S_j = 0 \tag{1}$$

It means that for finite or countably infinite sequence of disjoint sets, the length of the union of these sets is equal to the sum of the lengths of these sets.

The function ϕ is countably sub-additive when

$$\phi(S) \leq \sum_{i=1}^{\infty} \phi(S_i) \quad where \;\; S \in \mathbb{S}\,; \;\; S \subseteq \bigcup_{i=1}^{\infty} S_i \tag{2}$$

2.6 Metric Space

A metric space is a set wherein the concept of length or distance between elements is defined. It is denoted by (S, d) where S is a set and d is the metric (a single-valued, non-negative distance function) such that $d : S \times S \to R$. It satisfies the properties of positivity, symmetry, identity and triangle inequality.

If every Cauchy sequence in a metric space (S, d) converges to an element in the space, then the space (S, d) is said to be *complete*, e.g. The open interval $(0, 1)$ is not complete whereas the closed interval $[0, 1]$ is complete. Every compact (closed and bounded) metric space is complete.

Every normed space (distance function set to the absolute difference between two elements of the space) is also a metric space. As a side note, a complete normed space is siad to be a *Banach space*.

3 Measure Space and Measure

If there is a countably additive set function μ defined on σ-algebra $\mathbb{S}$ of subsets of S , then the triplet $(S, \mathbb{S}, \mu)$ is a *measure space*. An example of measure space is an Euclidean space with Lebesgue measure. The sets of $\mathbb{S}$ are called *measurable sets* and the function μ is called a *measure* and has the following properties:

- μ is non-negative, i.e. $\mu(X) \geq 0$ for all $X \in S$.
- μ is countably additive (as defined in section 2.5).
- $\mu(\emptyset) = 0$.
- μ obeys monotonicity : $\mu(X) \leq \mu(Y)$ for all $X, Y \in S$ and $X \subset Y$.

The concept of *measure* has connections with *integration* over arbitrary sets. Moreover, if $\mu(S) = 1$, then the measure space is called a *probability space* and μ is a probability measure. The sets are called events. One says that a property holds *almost everywhere* if the set for which the property does not hold is a null set or a set with measure 0. In probability theory, analogous to *almost everywhere, almost certain* or *almost sure* means *except for an event of probability measure* 0. This means that the event has zero probability of not occurring although it is still possible that they might not occur. There are different kinds of measures - Borel measure, Lebesgue measure, Haar measure, Dirac measure to name a few.

The measure space $(S, \mathbb{S}, \mu)$ is complete if every subset of a set of measure 0 is measurable and has measure 0. Formally, for all $S \subset \mathbb{S}$, if $\mu(S) = 0$ then for all subsets $X \subset S$, $\mu(X) = 0$. If not, then a space can be made complete. The Borel measure is not complete and hence the complete Lebesgue measure is preferred. Every Borel set is also Lebesgue measurable.

3.1 Lebesgue Measure

The definition of measure as given above is of a generalized form. However, in practice it is often the case that only those sets that are Lebesgue measurable need to be considered. To get to that, let's define an outer measure $\mu^*(X)$ and an inner measure $\mu_*(X)$ on a set $X \subseteq S$.

The *outer measure* of a set $X \subseteq S$ is defined as:

$$\mu^*(X) = inf \left\{ \sum_n m(B_n) : m \; is \; a \; measure \; on \; interval \; B_n \; and \; X \subseteq \bigcup_n B_n \right\} \tag{3}$$

The outer measure as defined above is again of generalized form. Consider an interval B with endpoints a and b such that $a < b$ instead of the set A, then the outer measure for the interval is defined as $b - a$.

The *inner measure* is defined as :

$$\mu_*(X) = m(S) - \mu^*(S \setminus X) \tag{4}$$

If the following is true,

$$\mu^*(X) = \mu_*(X) \tag{5}$$

then $\mu(X)$ (which denotes the value of $\mu^*(X)$ or $\mu_*(X)$) is said to be a *Lebesgue measure* on the set X.

3.2 Lebesgue Integral

A measure μ of a measure space $(S, \mathbb{S}, \mu)$ is determined by its values $\mu(s_i)$ where $s_.$ is a finite or infinite sequence in the countable space S. Here, $\mathbb{S} = 2^S$. If f is a function from this countable space into $\mathbb{R}$ and is μ-integrable over the set S (i.e. the sequence evaluated under f is absolutely convergent), then the Lebesgue integral of f on S is defined as a sum of the sequence evaluated under f as shown below :

$$\int_S f(s) d\mu = \sum f(s_.)\mu(s_.) \tag{6}$$

If $\mu(S) = 1$ i.e. the σ-algebra $\mathbb{S}$ is a probability space and μ is a probability measure, then in probability theory, Equation 6 is referred to as the *expectation* of f and is denoted by $E(f)$.

A generalized definition of the integral of f when there exists a sequence of functions $\{f_n\}$ μ-integrable over S and uniformly convergent to f is given by:

$$\int_S f(s) d\mu = \lim_{n \to \infty} \int_S f_n(s) d\mu \tag{7}$$

The definition of an integral in this form is in contrast with that of the Riemann integral which is only applicable to intervals and unions of intervals. For Riemann integral (following the idea of method of exhaustion), the space over which f is defined is divided into smaller sub-intervals and represented in the following way :

$$\sum_{n} f(\eta_n)(x_{n+1} - x_n) \tag{8}$$

where the value of $f(x)$ is substituted by the function evaluated at an arbitrarily chosen point η within the interval $[x_n, x_{n+1}]$. However, this can be done only when $f(x)$ is continuous or when it does not have many points of discontinuity. The Riemann integral of f is obtained by taking the limit of n to infinity in the above sum.

In order to distinguish between the Lebesgue and Reimann integrals consider the values that the function f can take to be on the x-axis (called the domain of the function) and the values of the function evaluated at the chosen points to be on the y-axis (called the range of the function). The Lebesgue integral defines the sub-intervals along the range of the function and the Riemann integral defines them along the domain of the function. Thus, the Riemann integral requires the values of the domain to be very close to each other (function should be continuous). This is a limitation since it restricts the class of functions on which the integral can be defined. Thus the Lebesgue integral is preferred because it is applicable to arbitrary classes of functions. Another reason is that if there is a sequence of Riemann integrable functions $\{f_n\}$, then it is not true that the function f defined as $f = \lim_{n\to\infty} f_n$ (absolutely convergent) is Riemann integrable.

Now, I present two theorems from Kolmogorov and Fomin [1999] that explicitly state the conditions under which the limit of an integral is equal to the integral of the limit as given in the definition in Equation 7. The need to take the limit inside the integral or to consider piecewise integration of a convergent sequence of functions arises in serveral applications, most notable being Fourier series.

Theorem 3.1 ***(Dominated Convergence Theorem).*** *If $\{f_n\}$ is a sequence of measurable functions from a measure space $(S, \mathbb{S}, \mu)$ into $\mathbb{R}$ and the sequence converges to f and*

$$|f_n(s)| \leq g(s)$$

almost everywhere, for all n, where g is a non-negative function integrable over S, then the limit function f is integrable over S and

$$\int_S f(s)d\mu = \lim_{n\to\infty} \int_S f_n(s)d\mu$$

Theorem 3.2 ***(Monotone Convergence Theorem).*** *If $\{f_n\}$ is a sequence of measurable functions from a measure space $(S, \mathbb{S}, \mu)$ into $\mathbb{R}$ such that $f_1 \leq f_2 \leq \ldots \leq f_n \leq \ldots$ and the sequence converges to f, then the limit function f is integrable over S and*

$$\int_S f(s)d\mu = \lim_{n\to\infty} \int_S f_n(s)d\mu$$

The theorem 3.1 is so called because of the use of the integrable function g for stating the condition of integrability of piecewise functions $\{f_n\}$ and the theorem 3.1 is so called because it assumes that the sequence of piecewise functions is monotonically increasing.

4 Measurable Function

It has been proven that it is impossible to assign a measure to all subsets of $\mathbb{R}$ without preserving some of the properties like additivity. Hence, as seen in the previous section I will concentrate only on certain *measurable* sets. Consequently, I will generalize the concept of a *function* adapted to the class of measurable sets and hence to their corresponding σ-algebras. Such functions are called *measurable functions*. They are defined as follows: Let $(S, 2^{S^{'}})$ be a countable measurable space and consider functions from a measurable space $(S, \mathbb{S})$ into $S^{'}$. Such a function f is measurable if the following is true for every set $X \subseteq S^{'}$:

$$f^{-1}[X] = \{s \in S \,|\, f(s) \in X\} \tag{9}$$

Equation 9 is also called the preimage (not an inverse function) of set X under the measurable function f. If $(S, \mathbb{S})$ is equipped with a probability measure, a measurable function is a *random variable* in the context of probability theory.

5 Final Comment

Measure theory is a fascinating area of study in its own right and at the same time plays an important role in functional analysis, advanced probability theory and statistics. In this note, I have only laid down a framework that gives an intuitive insight into the world of measures and enables us to define integrals, probability distributions and random variables. For more rigorous proofs, convergence properties and for exploring the use of Lebesgue integral in the context of square integrable functions and Hilbert spaces, the reader is referred to Kolmogorov and Fomin [1999].

A Few definitions

A.1 Power Set

A power set of a set S is the set of all possible subsets of S. It is denoted by 2^S or $\mathcal{P}^S$. If there are n elements in the set S, then $|\mathcal{P}^S| = 2^n$. 2^S defines the set of all functions from S to $\{0, 1\}$. A power set forms an abelian group when considered with the operation of symmetric difference and a commutative ring when considered with the operations of symmetric difference and intersection.

A.2 Convex Set, Convex Closure, Simplex

A set X in the linear space S is said to be *convex* if given any two arbitrary points x and y from the set, the segment joining them also belongs to X.

The *convex closure* of the set X is the smallest closed convex set containing X. The convex closure can also be obtained as the intersection of all the closed convex set which contain the set X.

A set of points $x_1, x_2, \ldots x_{n+1}$ in a normed linear space is said to be in *general position* if no $k+1 (k < n)$ of these points lie in a subspace of dimension less than k (if it does, then it is the *degenerate* case). The convex closure of such a set of points $x_1, x_2, \ldots x_{n+1}$ that are in general position is called an *n-dimensional simplex*. In other words, it is the minimal convex set. The n-dimensional simplex with $n + 1$ points is also represented in the following form:

$$x = \sum_{k=1}^{n+1} \alpha_k x_k \; ; \quad \alpha_k \geq 0 \quad and \quad \sum_{k=1}^{n+1} \alpha_k = 1 \tag{10}$$

A.3 Inequalities

Let X be a measurable space with dimension n; $a_1, a_2, \ldots a_n$ and $b_1, b_2, \ldots b_n$ are real or complex numbers; $1 \leq p$, $q \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$.

A.3.1 Cauchy-Schwarz Inequality

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 \leq \sum_{k=1}^{n} a_k^2 \sum_{k=1}^{n} b_k^2 \tag{11}$$

The Cauchy-Schwarz inequality is used to prove triangle inequality.

A.3.2 Hölder's Inequality

$$\sum_{k=1}^{n} |a_k b_k| \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} |b_k|^q\right)^{\frac{1}{q}} \tag{12}$$

Hölders inequality is a generalized case of Cauchy-Schwarz inequality and is used to prove the generalization of triangle inequality and the Minkowskis inequality.

A.3.3 Minkowski's Inequality

$$\left(\sum_{k=1}^{n} |a_k + b_k|^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} |b_k|^q\right)^{\frac{1}{q}} \tag{13}$$

A.3.4 Weak and Strong Convergence

A sequence $\{x_n\}$ in a normed linear space L is said to *converge weakly* to an element x when the following conditions are true:

- The norms of the elements $\{x_n\}$ are uniformly bounded : $||x_n|| \leq M$.
- $f(x_n) \to f(x)$ for all $f \in \bar{L}$

The sequence $\{x_n\}$ is said to *converge strongly* when its norm converges to x such that $||x_n - x|| \to 0$ as $n \to \infty$. Obviously, for finite dimensional spaces, strong convergence and weak convergence is the same. We have analogous definitions for convergence of linear functionals.

References

A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*, volume I and II. Dover Publications, Mineola, NY, 1999. ISBN 0-486-40683-0.