Measure Theory

Bill Last Updated:

22 March, 2020

ii

Contents

Motivation

Based on Dr. Dennis Cox's "The Theory of Statistics and Its Applications," I made the note for personal study purpose using RMarkdown and being posted in Github. All copyright belongs to Dr. Cox, and thus please do not copy, download, or distribute under any circumstance.

Chapter 1

Measure Spaces

1.1 Measure

A measure is a function μ defined for certain subsets A of a set Ω which assigns a nonnegative number $\mu(A)$ to each "measurable" set A. In probability theory, the probability is a measure, denoted by P instead of μ, which satisfies $P(\Omega) = 1$. In the context of probablity theory, the subset A is called an event, and Ω is called the sample space.

Example

(1) A die with 6 faces. Thus, the sample space Ω is the finite set of integers $\{1, 2, 3, 4, 5, 6\}$ corresponding to the possible outcomes if we roll the die once and count the number of spots on the fact that turns up. We define a probablity measure P hese events by $P(A) = Count(A)/6$.

(2) Let a random number be chosen in the interval $[0, 1]$ such that the probability of the number lying in any subinterval $[a, b](0 \leq a < b \leq 1)$ is the length (i.ee., $P([a, b]) = b - a$). Thus, here $\Omega = [0, 1]$. Such a random number is said to be uniformly distributed on the interval $[0, 1]$. We can extend the probability measure P from closed intervals to other subsets of $[0, 1]$ (e.g., $P((a, b)) = P([a, b)) = P((a, b]) = P([a, b]) = b - a$). Also, if $[a_1, b_1], [a_2, b_2], ...$ is a finite or infinite sequence of disjoint closed intervals (one can also allow open or semi-open intervals), then we can get $P(\cup_i [a_i, b_i]) = \sum_i (b_i - a_i)$.(It turns out for technical reasons that in this case, one cannot define probability measure of all subsets of $[0, 1]$.)

The probability measure of this example is related to a (nonprobability) measure: Let $\mu = m$ be a measure on arbitrary intervals of real numbers which equals the length of the interval, i.e., $m((a, b)) = b - a$ for any open interval

$(a, b), a < b$, and similarly for the other varieties of intervals. Here, $\Omega = \mathbb{R}$, the set of all real numbers, also denoted $(-\infty, \infty)$. This measure m is called Lebesgue measure.

1.1.1 σ-Fields

Definition 1

Let $\mathcal{F}$ be a collection of subsets of a set Ω. Then, $\mathcal{F}$ is called a sigma field (or, sigma algebra; written as σ-field or σ-algebra) if and only if it satisfies the following properties:

(i) The empty set $\emptyset \in \mathcal{F}$.

(ii) If $A \in \mathcal{F}$, then the complement $A^c \in \mathcal{F}$.

(iii) If $A_1, A_2, \dots$ is a sof elements of $\mathcal{F}$, then their union $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A pair $(\Omega, \mathcal{F})$ consisting of a set Ω and a $\sigma - field$ of subsets $\mathcal{F}$ is called a measurable space. The elements of $\mathcal{F}$ are called measurable sets or events.

Remarks 1

(1) The set Ω is called sample space in probability, but in general measure theory it is called the underlying set or underlying space.

(2) Since $\emptyset^c = \Omega$, it follows (i) and (ii) $\Omega \in \mathcal{F}$.

(3) Given any set Ω, the trivial σ-field is $\mathcal{F} = \{\emptyset, \Omega\}$. One can easily verify that this is a σ-field, and is in fact the smallest σ-field on Ω.

(4) Given any set Ω, the power set:

$$\mathcal{P}(\Omega) = \{A : A \in \Omega\}$$

Consisting of all subsets of Ω is also a σ-field on Ω, and in fact is the largest σ-field on Ω (Note: in some text, people denote $\mathcal{P}(\Omega)$ by 2^{Ω}).

(5) It follows from the definition that if $\mathcal{F}$ is a σ-field and $A_1, A_1, \dots$ is a sequence in $\mathcal{F}$, then the intersection $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Remarks 2

(1) A set A is called countable it can be listed as a sequence, finite, or infinite:

$$A = \{a_1, a_2, ...\}$$

We shall sometimes say that a set is countably infito indicate that it is countable but not finite.

(2) Any set which can be put in one to one correspondence with a subset of the natural numbers or "counting numbers" $\mathcal{N} = 1, 2, 3...$ is called countable.

(3) The set of all real numbers $\mathbb{R}$ (including irrational numbers like $\sqrt{2}$, π, e) can not put into a one to one correspondence with the natural numbers, so it has "more" elements, and is said to be uncountably infinite. These issues are pertinent to the technical difficulties wich make it impossible to extend Lebesgue measure to all subsets of $\mathbb{R}$, and hence require us to consider the notion of a σ-field (rather than just defining a measure on all subsets of the underlying space).

It takes a certain amount of work to obtain σ-field other than the trivial σ-field (i.e., $\{\emptyset, \Omega\}$) or the power set (i.e., $\mathcal{P}(\Omega) = \{A : A \in \Omega\}$). A standard approach is to consider the smallest σ-field containing a given family of sets. We shall illusrtate this concept. Let $A \in \Omega$ be a nonempty proper subset of Ω (i.e., $\emptyset \neq A \neq \Omega$), then

$$\sigma(A) = \{\emptyset, A, A^c, \Omega\}$$

is a σ-algebra (or, σ-field).

Example

If $\emptyset \neq B \neq \Omega$, $A \cap B \neq \emptyset$, and neither A or B is a subset of the other, then one can obtain a σ-fied $\sigma(\{A, B\})$ consisting of 16 element (i.e., $2^4 = 16$).

Definition 2

If $\mathcal{C}$ is any collection of subsets of Ω, then the σ-field generated by $\mathcal{C}$, denoted $\sigma(\mathcal{C})$, is the smallest σ-field containing $\mathcal{C}$. Hence, "smallest" means that if $mathcalF$ is any σ-field containing $\mathcal{C}$, then $\sigma(\mathcal{C}) \in \mathcal{F}$.

Definition 3

The Borel σ-field $\mathcal{B}$ on $\mathbb{R}$ is the σ-field generated by the collection of all finite open intervals. In symbols,

$$\mathcal{B} = \sigma((a, b) : -\infty < a < b < \infty)$$

The elements of $\mathcal{B}$ are called Borel sets.

1.1.2   Measures: Formal Definition

Definition 4

A measure space $(\Omega, \mathcal{F}, \mu)$ is a triple, consisting of an underlying set Ω, a σ-field $\mathcal{F}$, and a function μ, which is called "the measure" with domain $\mathcal{F}$ and satisfying the following:

(i) $0 \leq \mu(A) \leq \infty$ for all $A \in \mathcal{F}$.

(ii) $\mu(\emptyset) = 0$

(iii) If $A_1, A_2, \ldots$ is a sequence of disjoint elements of $\mathcal{F}$ (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then,

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$

A probability space is a measure space $(\Omega, \mathcal{F}, P)$ for which $P(\Omega) = 1$. (note, it replaces μ with P). A measure on $(\mathbb{R}, \mathcal{B})$ is called a Borel measure.

In probability theory, disjoint events are usually called mutually exclusive.

Remarks 3

Note that $\mu(A) = \infty$ is possible (unless $A = \emptyset$). In order for (iii) to make sure, we must know how to do arithmetic with ∞. For now, it suffices to know

$$\infty + x = \infty \; for \; x \in \mathbb{R}$$

$$\infty + \infty = \infty$$

Thus, if in (iii) in the Definition 4, $\mu(A_i) = \infty$ for some j, then $\sum_i \mu(A_i) = \infty$. Of course, $\sum_i \mu(A_i)$ may equal to ∞ even if all $\mu(A_i) < \infty$.

We will also need to know some multiplication rules for ∞ as well:

$$a \cdot \infty = \infty, \; for \; all \; a > 0$$

$$0 \cdot \infty = 0$$

Example 1.1.3

Let $(\Omega, \mathcal{F})$ be any measurable space. Let $A \in \mathcal{F}$ and define th measure #(A)=the number of elements of A. If A is an infinite set, then of course #(A) $= \infty$. It is fairly easy to check that $(\Omega, \mathcal{F}, \#)$ is a measure space, based upon the Definition 4.

Example 1.1.4

A special kind of measure is often useful is the unit point mass at x. Given a measurable space $(\Omega, \mathcal{F})$ and $x \in \Omega$, put

$$\delta_x(A) = \begin{cases} 1 & if \ x \in A \\ 0 & if x \notin A \end{cases}$$

It is easy to check that δ_x is a probability measure on $(\Omega, \mathcal{F})$.

Note that counting measure on $\{x_1, x_2, ...\}$ may be written in terms of unit point masses as follows. Thus, the sum o fmeasures is a measure.

$$\# = \sum_i \delta_{x_i}$$

Theorem 1.1.3 (Existence of Lebesgue Measure)

There is a unique Borel measure m satisfying

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

for every finite closed interval $[a, b]$, $-\infty < a < b < \infty$

Thinking of Lebesgue measure as a "mass distribution", we see that it is a continuous distribution of mass which is "concentrated" on a line (the real line) and assigns one unit of mass per length.

Proposition 1.1.4 (Basic Properties of Measures)

Within a measure space $(\Omega, \mathcal{F}, \mu)$, we can get

(a) Monotonicity: $A \in B$ implies $\mu(A) \le \mu(B)$, assuming A and B are in $\mathcal{F}$.

(b) Subadditivity: If $A_1, A_2, ...$ is any sequence of measurable sets, then

$$\mu(\cup A_i) \le \sum \mu(A_i)$$

(c) If $A_1, A_2, ...$ is a decreasing sequence of measurable sets (i.e., $A_1 \supset A_2 \subset ...$), and if $\mu(A_1) < \infty$, then

$$\mu(\cap_{i=1}^{\infty} A_i) = lim_{i \to \infty} \mu(A_i)$$

Proposition 1.1.5

(a) Let $\mu_1, \mu_2, ...$ be a finite or infinite sequence of measures on $(\Omega, \mathcal{F})$. Suppose $a_1, a_2, ...$ are nonnegative real numbers. Then $\mu = \sum_i a_i \mu_i$ is a measure on $(\Omega, \mathcal{F})$.

(b) Consider the same setup as in part (a). If each of the μ_i is a probability measure and if $\sum_i a_i = 1$, then μ is a probability measure.

(c) Let μ be a measrue on $(\Omega, \mathcal{F})$and let $A \in \mathcal{F}$. Define $v(B) = \mu(B \cap A)$ for all $B \in \mathcal{F}$. Then, v is a measure on $(\Omega, \mathcal{F})$.

1.1.3 Distribution Functions

If P is a probability measure ($p.m.$) on $(\mathbb{R}, \mathcal{B})$, i.e., a Borel $p.m.$, then we can define

$$F(x) = P((-\infty, x])$$

for all $-\infty < x < \infty$. F is called the (cumulative) distribution fucntion (c.d.f.) for P.

Note: Sometimes a probability measure itself is referred t obe a distribution. In this text, we will never use "distribution" without the word "function" to refer to $c.d.f.$ Note that the probability measure and the $c.d.c.$ are two different kinds of objects. One maps Borel sets to real numbers and the ohter maps real numbers to real numbers.

Theorem 1.1.5