# Measure Theoretic Foundations of Probability

Matt Cocci

February 8, 2013

## 1 Introduction

The following distillation pulls together several resources, including many Wikipedia articles and my course text, to try to succinctly describe the set- and measure-theoretic foundations of arbitrary probability spaces.

## 2 Sample Space

A *sample space*, defined $\Omega$, is an exhaustive set of all 'basic' outcomes that can occur. The word 'outcome'—in contrast to 'event'—specifically relates to these most fundamental occurrences, such as the different faces of a die or the different cards in a standard deck.[1] Only one outcome or $\omega \in \Omega$ can occur when we make an observation. So that if we know $\omega_0 \in \Omega$ happened, then no other $\omega_i \in \Omega$ can have occured, where $\omega_i \neq \omega_0$. Note that we may alternatively call outcomes—those $\omega \in \Omega$—by the more standard name of *sample points*.

To take an example, if we look at the roll of a die, the sample space will simply be

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

An observation will simply be whatever face turns up on a given roll.

## 3 The Notion of a $\sigma$-algebra

Once we have a sample space, $\Omega$, we can go a bit further. Typically, one sample point in isolation is not very interesting, except in a few rare instances—and most often with small samples spaces. But for larger sample spaces like the positive integers or the real

[1]Note that the terminology 'outcome' versus 'event' (which we will develop below) is my own.

line, the notion of events and $\sigma$-algebras will come in handy.

Let's start with events. An *event* is a any subset of, $E \subset \Omega$. It is effectively a collection of outcomes, like the set of all the outcomes where a die roll yields an odd number. Note that unlike *outcomes*, or $\omega \in \Omega$, *events* can overlap each other, and more than one can occur given an observation or basic outcome. For example, suppose we throw a one; then the following events have occurred: we threw a one, we threw a prime, we threw a number less than four, and so on ad infinitum.

Next, we define a $\sigma$-**algebra**, which I'll denote as $\mathcal{F}$ to be a collection of subsets of the sample space, a collection of events, which exhibits the following properties:

i. $\Omega \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$. Intuitively, it ensures something will happen.

ii. If the set $B \subset \Omega$ is in $\mathcal{F}$, then so is it's complement, i.e. $\Omega \setminus B \in \mathcal{F}$.

iii. Countable unions and intersections of elements of $\mathcal{F}$ are also in the $\mathcal{F}$. Mathematically,

$$B_i \in \mathcal{F}\ \forall i \ \Rightarrow \bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$$

$$B_i \in \mathcal{F}\ \forall i \ \Rightarrow \bigcap_{i=1}^{\infty} B_i \in \mathcal{F}$$

Any set $X \in \mathcal{F}$ is called a **measurable set**.[2] Note by the way we defined a $\sigma$-algebra, the measurable sets, or events, in $\mathcal{F}$ will fail to be disjoint.

## 3.1 Simplest Example

For an even simpler example than die rolling of something that is a perfectly valid $\sigma$-algebra, we sometimes take the power set of $\Omega$—the collection of all possible subsets of $\Omega$—to be our $\sigma$-algebra. It will surely satisfy the necessary properties, and it happens to be very easy to denote.

## 3.2 Big Picture

One of the reasons why we employ two levels of complexity in having *both* a sample space, $\Omega$, and a $\sigma$-algebra, $\mathcal{F}$, is that it allows us to properly make sense of complements of sets. Specifically, if $A \in \mathcal{F}$ represents rolling a 1, it's a natural thing in probability to want to talk about the chances of *not* rolling a 1—the complement of $A$. The preceding definitions make that a little more tractable.

[2] Just as Topology is based on open sets, which are members of some $\sigma$-algebra, so too is Measure Theory baed on measurable sets which are members of some $\sigma$-algebra.

Specifically, if we try to talk about $A$'s complement, that only makes sense if $\mathcal{F}$ is built upon the foundation of a more "meta" set. [3] For example, if we want to talk about the complement, $A^C$, without specifying that "meta" set, we could technically say that aliens landing on Earth tomorrow is in the complement of $A$—for it's definately not the event where you roll a 1. But it's not very relevant, and our upcoming definition of probability would force us to attach a probability to this as well—something not as easily quantified as rolling a die.

Therefore using a "meta" set, called the sample space $\Omega$ reduces the class of outcomes or events to what's relevant. We can just specify $\Omega$ to be the outcomes from rolling a die, and we rid ourselves of many headaches, additional considerations, and inter-planetary considerations.

# 4 Measure

Once we have a $\sigma$-algebra, we can pair the sample space with it's $\sigma$-algebra to form what's called a **measurable space**, which is simply the ordered pair $(\Omega, \mathcal{F})$. From there, we can start assigning probabilities to those elements in the $\sigma$-algebra which all have some chance of occurring. To do so, we will use a special type of function, called a measure.

A **measure** is a function that maps sets into real numbers. In our case, we will define a specific type of measure, a **probability measure**, as a function that has as it's domain a $\sigma$-algebra and maps measurable sets in the $\sigma$-algebra into the real line, subject to a few conditions. More precisely, it has the following characteristics:

i. If $P$ is our probability measure and $\mathcal{F}$ is our $\sigma$-algebra, then

$$P : \mathcal{F} \to \mathbb{R}$$

ii. $P(F) \in [0, 1]$ for all measurable sets $F \in \mathcal{F}$. In addition, $P(\Omega) = 1$.

iii. Our probability measure satisfies the probability axioms. And if $F_i$ are all disjoint and members of the $\sigma$-algebra, then

$$p(\cup F_i) = \sum P(F_i)$$

**Definition** Now that we have our definition of a probability, I just want to define a common term more precisely. Specifically, if a property is true *except* for an event of probability 0, then we say that property holds "almost surely."

[3] I use "meta" to avoid using a term like "higher" or "bigger" because, typically, the $\sigma$-algebra will be bigger than or at least as big as $\Omega$ in cardinality, and I don't want to cause any confusion. And in addition, I wanted to talk about using the word "meta" so that I can reasonably call this footnote "meta-meta."

# 5 Probability Space

So now we have a way of assigning probabilities to any arbitrary event in our measurable space, $(\Omega, \mathcal{F})$. If we join the probability measure, $P$, to our measurable space, we obtain a proper **Probability Space**. It is defined as the ordered triplet

$$(\Omega, \mathcal{F}, P)$$

# 6 Measurable Function and Random Variables

One of the fundamental notions of probability is that of a random variable; therefore, making the definition more rigorous deserves some attention.

So let's start with the notion of a **measurable** function. If $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two measurable spaces then, a function

$$f : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$$

is a measurable function if

$$f^{-1}(B) \in \mathcal{F}_1, \ \ \forall B \in \mathcal{F}_2$$

In words, a function is measurable if the pre-image of an element in the $\sigma$-algebra of the co-domain is in the $\sigma$-algebra of the domain. Simply, it preserves the structure between two measurable spaces, sending measurable sets in one measurable space to measurable sets in the other.

Turning to probability, suppose that we have our probability space, $(\Omega_1, \mathcal{F}_1, P)$, and another measurable space, $(\Omega_2, \mathcal{F}_2)$. Then an $\boldsymbol{(\Omega_2, \mathcal{F}_2)}$**-valued random variable** is a *measurable function*

$$X : (\Omega_1, \mathcal{F}_1) \to (\mathbb{R}, \mathcal{F}_2)$$

Since $X$ is a measurable function, we know that

$$X^{-1}(B) \in \mathcal{F}_1, \ \ \forall B \in \mathcal{F}_2$$

in which case we say that $X$ is $\mathcal{F}_1$-measurable. To clarify further

$$X^{-1}(B) = \{\omega \in \Omega_1 | X(\omega) \in B\}$$

**Note** Oftentimes, we'll just take the set $\mathcal{F}_2$ to be some properly defined $\sigma$-algebra on the real line, like the Borel Set (defined below).

## 6.1 Sigma Algebras Generated by Random Variables

Above, we assumed that we knew the $\sigma$-algebra for $\Omega_1$ already—that it was given. But we could just as well *generate* one from some random variable $X$, and this generated $\sigma$-algebra could very well differ from $\mathcal{F}_1$.

So let's assume that $X$ is a random variable

$$X : (\Omega_1, \mathcal{F}_1) \to (\mathbb{R}, \mathcal{F}_2).$$

Then the $\sigma$-algebra generated by $X$, denoted by $\sigma(X)$, is defined as *all* sets of the form

$$\sigma(X) := \{\omega \in \Omega_1 | X(\omega) \in A\}, \ \ \forall\, A \subset \mathbb{R}$$

which can be written more compactly as

$$\sigma(X) = \{X^{-1}(A) | A \subset \mathbb{R}\}$$

**Definition** Finally, if $\mathcal{G}$ is a sub-$\sigma$-algebra of the set $\mathcal{F}$ defined above, then the random variable $X$ is $\mathcal{G}$-measurable if

$$\sigma(X) \subset \mathcal{G}$$

Let's give a concrete example. Suppose $\Omega$ consists of all the possible combination of up and down moves in a binomial tree. $X$ is a random variable denoting stock price. Then for a set of real numbers representing the possible prices the stock could take on (this is our $A \subset \mathbb{R}$), the set $\sigma(X)$ will be the sigma algebra resulting from the set of all possible paths that that the stock could have taken to get to those prices in $A$.

## 6.2 Random Variables and Their Distributions

A random variable is actually quite distinct from its distribution. Recall that a Random Variable is just a mapping from the sample space $\Omega$ into something more tractable, like the real line. Therefore, we could discuss this mapping–the random variable– without ever considering probabilities.

However, oftentimes we will want to discuss probabilities, but the sufficiently general definition of the random variable just given offers a lot of flexibility. So much, in fact, that a random variable could have more than one distribution, as we know occurs in stock process which have traditional probabilities and also *risk-neutral* probabilities (or *pseudo-probabilities*).

So let's define a **distribution** as a measure that assigns probabilities to the real numbers that a random variable generates (after being passed an element in the sample space). The most natural distribution is the induced measure, $\mathcal{L}_X$ defined as follows

$$\mathcal{L}_X(A) := P\{X \in A\}, \ \ A \subset \mathbb{R}$$

Let's unpack that. If $A$ is a subset of the real line, it the random variable may or may not take on values in that subset, so we want a notion of probability. Well, we can take the pre-image, $X^{-1}(A)$, and look at all those elements of the sample space that map to $A \subset \mathbb{R}$ under the random variable $X$. Then, once we have elements of the sample space, we can use our traditional notion of the $\sigma$-algebra, $\mathcal{F}$, and the probability measure $P$ that are already defined on the probability space to assign the probabilities.

Thus, we can associate probabilities with our random variable so long as there exists a distribution measure, like $\mathcal{L}_X$. But recall that $\mathcal{L}_X$ was not unique. It's simply a function to assign probabilities to values that the random variable can take on. We could consider other measures that also accomplish this, and that insight legitimizes the use of such things as *risk-neutral probabilities*.

# 7 Lebesgue Measure and the Lebesgue Integral

## 7.1 Introduction

The **Lebesgue Measure** is the standard way of assigning a measure to subsets of $n$-dimensional Euclidean space. If we restrict to $\mathbb{R}$, then the Lebesgue Measure of intervals is simply the length. But rather than consider all of $\mathbb{R}$, we'll restrict further to *Borel Sets*.

This will allow us to construct the *Lebesgue Integral*, a generalization of the Riemann Integral.

## 7.2 Borel Sets

The *Borel $\sigma$-algebra*, denoted $\mathcal{B}(\mathbb{R})$, is the smallest $\sigma$-algebra containing (and, in a sense, generated by) all open intervals in $\mathbb{R}$.

**Examples** The following are a few samples of Borel Sets in $\mathcal{B}(\mathbb{R})$:

- Every open interval of the form $(a, b)$.
- The open rays $(a, +\infty)$ and $(-\infty, b)$.
- All unions of the form
$$B_1 \cup B_2, \quad B_1, B_2 \in \mathcal{B}(\mathbb{R})$$
- All complements of sets in $\mathcal{B}(\mathbb{R})$, since it's a sigma algebra. Note that this implies all *closed* intervals in $\mathbb{R}$ are Borel Sets as well, in addition to all half-open and half-closed intervals.
- All one point sets, $a \in \mathbb{R}$, as we see that it is in the intersection of other Borel Sets, implying inclusion in $\mathcal{B}(\mathbb{R})$ since it's a $\sigma$-algebra
$$\{a\} = \bigcap_{n=1}^{\infty} \left( a - \frac{1}{n}, a + \frac{1}{n} \right).$$
- The last item implies that all finite countable collections of points in $\mathbb{R}$ are Borel Sets too. Therefore, the set of all rational numbers is Borel since coutable, and the set of all irrational numbers is Borel since it's the complement of a set in the sigma algebra.

However, it should be noted that not all sets of real numbers are Borel Sets. In particular, any non-Borel set must be uncountable (though the opposite is not ture, as shown above).

## 7.3 Lebesgue Measure

Let's start more generally and define a *measure* on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to be a function

$$\mu : \mathcal{B} \to [0, \infty)$$

with the following properties

i. $\mu(\emptyset) = \emptyset$.

ii. If $A_1, A_2, \ldots$ are disjoing sets in $\mathcal{B}(\mathbb{R})$, then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \bigcup_{k=1}^{\infty} \mu(A_k).$$

We define the **Lebesgue Measure** on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to the measure $\mu_0$ that assigns the measure of each interval to be its length.

## 7.4 Functions in This World

Let $f$ be a function

$$f : \mathbb{R} \to \mathbb{R}.$$

We say that $f$ is **Borel-Measurable** if

$$A \in \mathcal{B}(\mathbb{R}) \Rightarrow f^{-1}(A) \in \mathcal{B}(\mathbb{R}).$$

Or equivalently, we could say that we want the $\sigma$-algebra generated by $f$ to be contained in $\mathcal{B}(\mathbb{R})$.

## 7.5 The Lebesgue Integral

Let $I$ be the *indicator function.* We define

$$A := \{x \in \mathbb{R} | I(x) = 1\}$$

to be the set *indicated by* $I$.

The **Lebesgue Integral** of $I$ is defined

$$\int_{\mathbb{R}} I d\mu_0 = \mu_0(A).$$