MEASURES

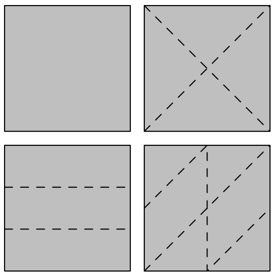
M

1. Fundamentals

A measure is a mathematical object that quantifies the size of sets. Each distinct measure embodies a different way to assess how big a set is. At first thought, it would seem that there is only one natural measure of a set's size—its cardinality. Indeed, the cardinality is a measure as we will define it below; given a set $\mathcal{A}$, $\#\mathcal{A}$ is a non-negative, possibly infinite, number. But if we examine the issue carefully, we realize that there are other reasonable ways to measure the size of a set. Consider the two sets $[1, 10]$ and $[10, 100]$. Both have the same cardinality ($= \#\mathbb{R}$), but our physical intuition suggests that the latter is "bigger" somehow. Indeed, we can compute the length of the two intervals, 9 and 90 respectively, which confirms our intuition. So, the length of a set provides a different measure of its size. We can also think about the lengths of the two intervals on a logarithmic scale (base 10): [1,10] spans the exponents 0 through 1 and [10,100] spans the exponents 1 through 2. Measured this way, the two sets have the same size, 1, which is finite.

Measures are based on a familiar idea—size. However, they may seem a bit abstract because mathematicians like to take a concept, whittle it down to its barest properties, and then explore the implications of those properties in contexts that are very different from the original idea. While this often turns out to be a profitable enterprise, it can seem daunting at first. Even though measures are likely unfamiliar to you as mathematics, try to keep in mind the strong intuition about size on which they rest as you proceed through the material. To build on this, we begin our exploration of measures with three examples that are familiar from everyday life: length, area, and volume.

1



1.1. Length, Area, and Volume

We can measure the length of a curve, the area of a surface, the volume of some object. (Eureka!) From a mathematical perspective, the curve, surface, and object are just sets of points in space, so length, area, and volume give us a way to measure the extent of one, two, and three-dimensional sets respectively. This is the fundamental idea of a measure: a rule that associates to sets a number that quantifies the size of the set. To get an intuitive grasp of what a measure is, we will use length, area, and volume as our models. We will find the properties that these satisfy, and then ask which of these properties are essential for thinking about the size of sets. This will lead us to the general idea of a measure.

Property #1. Non-negativity. Length, area, and volume can in principle be any positive value; we just need to find a long enough curve, a big enough area, or a large enough object. They can be zero as well: a point is a curve that goes nowhere, a line is a rectangle with no width, a plane is a solid with no depth. If one is willing to posit an infinite space to the universe, we can imagine curves, surfaces, and volumes that are infinite as well. But negative length, negative area, negative volume have no meaning.

Property #2. Additivity. We can measure the length of a path on a map along roads from Pittsburgh, through Philadelphia and Trenton, New Jersey, to Newark, New Jersey. If we cut that path into three non-overlapping pieces – say from Pittsburgh to Philadelphia, Philadelphia to Trenton, and Trenton to Newark – then the lengths of the three pieces combined must be the same as the length of the original path. If I divide a rectangular region into non-overlapping pieces,[1] the total area of the pieces must equal the total area of the original region. Moreover, this same fact is true no matter how I partition the region. The same is true for volume as well.

We call this property *additivity.* To divide a set $\mathcal{A}$ – the points on a curve, in the rectangular region, or in the object – into non-overlapping pieces is to find a collection of pairwise disjoint sets $\mathcal{A}_1, \mathcal{A}_2, \ldots$ such that $\mathcal{A} = \bigcup_i \mathcal{A}_i$. The $\mathcal{A}_i$'s here represent the pieces into which we divide the original set. To say that the union of the pieces equals the original set is to say that we have included all the pieces that we made. To say that the sets are disjoint is to say that the pieces do not overlap. The additivity property states that the length (or area or volume, as appropriate) of $\mathcal{A}$ must equal the sum of the lengths (or areas or volumes) of the $\mathcal{A}_i$'s.

In our physical reality, we can imagine dividing a curve or surface or object into a finite (though perhaps large) number of pieces. If the additivity property holds for any finite number of pieces, we call it *finite additivity.* Mathematically it will be more convenient to require a stronger property: we require that additivity hold for any countable[2]number of pieces. We call this *countable additivity.*

2 Recall that countable means finite or countably infinite.

Any measure of size for which the additivity property holds must have another important property – monotonicity. To understand monotonicity, consider the path from Pittsburgh to Newark above. If we go along that path until Philadelphia and then stop, the length of the subpath cannot be bigger than the entire path. More generally, if $\mathcal{A} \subset \mathcal{B}$, then the size of $\mathcal{A}$ must be no greater than the size of $\mathcal{B}$ because of additivity: $\mathcal{B}$ can be divided into two disjoint pieces, one of which is $\mathcal{A}$, $\mathcal{B} = \mathcal{A} \cup \text{compl}\,(\mathcal{A})$.

Property #3. Empty Set. In daily life, we rarely try to compute the length, area, or volume of the empty set. But to take these ideas into the mathematical realm, we need to come up with a sensible assignment that is consistent with our physical intuition. Since the magnitude of length, area, and volume derives from the extent of the curve, surface, or object, and since the empty set by definition has no points and thus no extent, it makes sense to take the length, area, and volume of the empty set as zero.

Property #4. Other Null Sets. The length of a point is 0. The area of a line, or indeed any curve, is 0. The volume of a plane, or any surface, is 0. The existence of these "null sets" tells us a lot about the nature of length, area, and volume. In particular, length is inherently one-dimensional, area two-dimensional, and volume three-dimensional; if we compute the size of a set of too low a dimension, we et 0.

Property #5. Translation Invariance. If I take a curve and shift it in space without changing its shape, its length does not change. Similarly for a region or object. Length, area, and volume are unchanging (invariant) under shifts in space (translations). If this weren't true, then walking around town could change your size.

3 An interval is a one-dimensional hyper-rectangle. Remember that $a < b$ is implicit in this notation.

Property #6. Hyper-Rectangles. An interval of the form $[a, b] \subset \mathbb{R}^3$ has length $b-a$. This is consistent with the fact that a point has length zero, because the point $\{a\}$ can be written as $[a, a]$. The additivity of length implies that the intervals of the form $]a, b[$, $]a, b]$, and $[a, b[$ all have the same length $b - a$. To see why, notice that the interval $[a, b]$ can be written as a union of pairwise disjoint intervals in three ways:

$$\begin{aligned} [a, b] &= [a, a] \cup]a, b[\cup [b, b] \\ &= [a, a] \cup]a, b] \\ &= [a, b[\cup [b, b] \ , \end{aligned}$$

Additivity tells us that for any such disjoint partition of $[a, b]$, the sum of the lengths of the pieces must be $b - a$. Since the boundaries $[a, a]$ and $[b, b]$ have length 0 (they are zero-dimensional sets), all four types of intervals have the same length. This still holds if a or b are infinite; $]a, \infty[$, $]-\infty, b[$, and $]-\infty, \infty[$ all have length ∞.

Similarly, the area of a rectangle $[a_1, b_1] \times [a_2, b_2]$ is $(b_1 - a_1)(b_2 - a_2)$, and the volume of a rectangular solid $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$, $(b_1 - a_1)(b_2 - a_2)(b_3 - a_3)$. These hold for rectangles and rectangular solids with any part of the boundary missing because the boundary is a lower dimensional set.

These six properties determine length, area, and volume. In other words, any measure of set size that satisfies these properties must be length (if it is defined for one-dimensional sets), area (if it is defined for two-dimensional sets), or volume (if it is defined for three-dimensional sets). Our next task is to decide which of these six properties are quintessential properties for a measure of the size of a set and which are specific to length, area, and volume. Well, we only know one other candidate for a measure – cardinality – so finding which of these properties cardinality shares will help us determine the general features we wish a measure to satisfy.

For cardinality, only properties #1, #2, #3, and #5 hold. While these all seem quite general, it turns out that #5, translation invariance, is a bit too specialized. Here's a quick argument. If # is a measure of the size of sets, then the following should be as well: assign $\mathcal{A} \subset \mathbb{R}$ to be size $\#\mathcal{A} \cap [0, 1]$. This computes the cardinality of the part of $\mathcal{A}$

that lies in $[0, 1]$ and so has all of the features of cardinality except translation invariance. So we judge properties #1, #2, and #3 to be essential for measuring size.

This gives us a provisional definition of a measure: any function that maps a set to a non-negative (and possibly infinite) number, such that $\emptyset$ maps to 0 and the (countable) additivity property is satisfied. It turns out that there is nothing provisional here, as we will see below.

Before beginning with measures in the abstract, let's look a little more formally at the measures that correspond to length, area, and volume.

We begin in $\mathbb{R}$ and, since this is one-dimensional, with length. Suppose we have a subset $\mathcal{A} \subset \,]0, 1[$; how do we compute the length of $\mathcal{A}$? To make things interesting we will assume that $\mathcal{A}$ is a strict subset; we already know the length of $]0, 1[$.[4] What we are looking for is a mathematical algorithm that is guaranteed to produce the length of $\mathcal{A}$, although it may not be implementable in practice.

To start with, since $\mathcal{A} \subset \,]0, 1[$, we know that $\text{length}\Big(\mathcal{A}\Big) \leq 1$ by the monotonicity property of length mentioned above. Our approach will be to see how tight we can make this upper bound. Think about measuring the length of $\mathcal{A}$ by covering $\mathcal{A}$ with pieces of string. The subset $]0, 1[$ corresponds to a single piece of string of length 1; it covers $\mathcal{A}$ but if $\mathcal{A}$ has some gaps, our string is a bit too long. So, we can try to cover $\mathcal{A}$ by several pieces of string of total length < 1. Because $\mathcal{A}$ may have some very short parts, these pieces of string may be a little too long or may overlap each other a bit. That's ok for now. When we add up the length of all our pieces, we find that $\text{length}\Big(\mathcal{A}\Big) \leq$ the sum of the lengths of the pieces, which in turn is < 1. If we keep trying – cutting the pieces more and more carefully to get a tighter fit to $\mathcal{A}$ while still covering $\mathcal{A}$, then our upper bound on $\text{length}\Big(\mathcal{A}\Big)$ will get tighter and tighter. Thus, $\text{length}\Big(\mathcal{A}\Big)$ must be $\leq$ the total string length for every such covering of $\mathcal{A}$ by pieces of string, and we can get arbitrarily close to $\text{length}\Big(\mathcal{A}\Big)$ with some such covering.[5] It follows that $\text{length}\Big(\mathcal{A}\Big)$ must be the greatest lower bound of the lengths of all coverings of $\mathcal{A}$ by pieces of string.

Ok, enough with string. The pieces of string in the above argument are just analogies for intervals. Consider a countable collection of in-

4 Nonetheless, we need our algorithm to work for $]0, 1[$, so we will have to revisit this case.

5 In other words, there can be no number ℓ such that such that $\text{length}\Big(\mathcal{A}\Big) < \ell$ and $\ell <$ the string length of every covering of $\mathcal{A}$.

tervals $]a_i, b_i[$ that "cover" $\mathcal{A}$, that is

$$\mathcal{A} \subset \bigcup_i \,]a_i, b_i[\, .$$

We do not require the intervals to be disjoint.[6] We call such a collection of intervals a *covering of* $\mathcal{A}$. Since any overlaps will be counted multiple times when adding up the lengths of the intervals, the additivity property of length implies that

$$\text{length}\Big(\mathcal{A}\Big) \leq \sum_i \text{length}\Big(]a_i, b_i[\Big) = \sum_i (b_i - a_i).$$

The $\sum_i (b_i - a_i)$ is the "length of the covering".

We can then consider a series of tighter and tighter coverings, trimming a little overlap here, cutting a little excess there, making the length of the covering closer and closer to the length$\Big(\mathcal{A}\Big)$. We know two facts:

1. length$\Big(\mathcal{A}\Big) \leq$ the length of every covering,
2. We can find some covering whose length is arbitrarily close to length$\Big(\mathcal{A}\Big)$.

It follows that length$\Big(\mathcal{A}\Big)$ is the greatest lower bound of all the lengths of coverings of $\mathcal{A}$. We are done.

This algorithm defines a measure length$\Big(\Big)$ which satisfies all the properties alluded to above. We can check that it makes sense with two simple examples, $]0, 1[$ and a point $\{1/2\}$. In the former case, the simple covering $]0, 1[$ is perfect and gives us a length of 1. In the latter case, the intervals $]1/2 - 1/n, 1/2 + 1/n[$ for positive integer n each cover $\{1/2\}$, but since their length goes to 0 as n goes to ∞, length$\Big(\{1/2\}\Big)$ must be 0, as we expect.

The same algorithm works to define area in $\mathbb{R}^2$, except we use two-dimensional intervals (i.e., rectangles), and to define volume in $\mathbb{R}^3$, except we use three-dimensional intervals (i.e., rectangular solids). In fact, we can use the same argument in any $\mathbb{R}^k$ using k-dimensional hyper-rectangles. This means that length, area, and volume are intimately related (as we knew from the start) and differ only in dimensionality. We call this family of measures on $\mathbb{R}^k$ *Lebesgue measures*[7] after the French mathematician Henri Lebesgue.

6 If the intervals were disjoint, then the sum of their lengths would equal the length of $\mathcal{A}$ by countable additivity. The point of the algorithm is that it is harder to find disjoint intervals that cover $\mathcal{A}$ than it is to find intervals that cover $\mathcal{A}$ without this constraint.

7 Lebesgue is pronounced Leb-ache.

Lebesgue measure on $\mathbb{R}$ is length. Lebesgue measure on $\mathbb{R}^2$ is area. Lebesgue measure on $\mathbb{R}^2$ is volume. Lebesgue measure on $\mathbb{R}^k$ for $k > 2$ is a "hyper-volume", a direct generalization of length, area, and volume which satisfies properties #1–#6 above (assigning measure $(b_1 - a_1) \cdots (b_k - a_k)$ to the hyper-rectangle $[a_1, b_1] \times \cdots \times [a_k, b_k]$). Because these measures have all the same properties except dimensionality, we use the same symbol, μ_{Leb}, to refer to all of them, and let the domain ($\mathbb{R}$, $\mathbb{R}^2$, etc.) be clear from context.

8 In a completely rigorous definition of a measure, we would have to be careful about specifying the set of subsets on which the measure is defined (i.e., the domain of μ). The reason is that if countable additivity holds it is possible in some cases to construct anomalous sets that cannot be measured at all. Thus, the domain of μ would sometimes need to be a strict subset of $2^{\mathcal{X}}$ that itself satisfies certain requirements. This is an important but very technical point that is largely irrelevant to our efforts here, so we ignore it in favor of the "practical" definition given.

1.2. Measures Defined

Let $\mathcal{X}$ be any set. A measure on $\mathcal{X}$ is a function μ that maps the set of subsets of $\mathcal{X}$ to $[0, \infty]$ that satisfies (i) $\mu(\emptyset) = 0$ and (ii) the following *countable additivity* property:

Definition. For any countable and pairwise disjoint collection of subsets of $\mathcal{X}$ $\mathcal{A}_1, \mathcal{A}_2, \ldots$,

$$\mu\left(\bigcup_i \mathcal{A}_i\right) = \sum_i \mu(\mathcal{A}_i).$$

Remember that μ is just a function,[9] albeit on a strange domain, so if $\mathcal{A} \subset \mathcal{X}$, $\mu(\mathcal{A})$ is the image of the "point" $\mathcal{A}$ in $2^{\mathcal{X}}$ under the function μ, which is just a number. We call $\mu(\mathcal{A})$ the "measure of the set $\mathcal{A}$" or "measure of $\mathcal{A}$" for short. Notice that according to the definition, the measure of $\mathcal{A}$ may be ∞.

9 As we have defined it, μ: $2^{\mathcal{X}} \to [0, \infty]$

The countable additivity property seems rather abstract as it is stated above, but it comes with a definite physical intuitition that you should keep in mind: If we take a set and divide it into non-overlapping pieces, in any way, then the measure (think "size") of all the pieces adds up to the total measure (think "size") of the original set. See figure 1.

We have seen that the familiar ideas of length, area, and volume are measures, corresponding to Lebesgue measure on $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$, but these are so familiar that they can obscure some of the power of the idea. So, let's look at some very different examples of measures.

Example 1. How many elements? A natural measure associates to each set its cardinality. But this is exactly $\#(\mathcal{A})$ as we defined it earlier. Let us check that this actually defines a measure. First, $\#(\emptyset) = 0$ by definition. Second, if we take a collection of disjoint sets, then the cardinality of the union of these sets is the sum of the cardinalities because the sets do not overlap. Cardinality gives one way to quantify the size of a set; notice how it differs from length for subsets. For example, all intervals in $\mathbb{R}$ have the same cardinality.

Example 2. A point mass at 0. Consider a measure $\delta_{\{0\}}$ on $\mathbb{R}$ defined to give measure 1 to any set that contains 0 and measure 0 to any set that does not. Mathematically, we write the definition

$$\delta_{\{0\}}(\mathcal{A}) = \#(\mathcal{A} \cap \{0\}) = \begin{cases} 1 & \text{if } 0 \in \mathcal{A} \\ 0 & \text{otherwise,} \end{cases}$$

for $\mathcal{A} \subset \mathbb{R}$. We can similarly define a point mass at any point x by $\delta_{\{x\}}(\mathcal{A}) = \#(\mathcal{A} \cap \{x\})$.

Example 3. Counting measure on the integers. Consider a measure $\delta_{\mathbb{Z}}$ that assigns to each set $\mathcal{A}$ the number of integers contained in $\mathcal{A}$. We define this by

$$\delta_{\mathbb{Z}}(\mathcal{A}) = \#(\mathcal{A} \cap \mathbb{Z}).$$

Counting measure can be defined for sets other than $\mathbb{Z}$ in an analogous way.

Example 4. A Geometric Measure. Suppose that $0 < r < 1$. Define a measure on $\mathbb{R}$ that assigns to a set $\mathcal{A}$ a geometrically weighted sum over non-negative integers in $\mathcal{A}$. Specifically,

$$\mu(\mathcal{A}) = \sum_{i \in \mathcal{A} \cap \mathbb{Z}_{\oplus}} r^i.$$

For example, it follows that $\mu(\mathbb{R}) = \mu(\mathbb{Z}) = \mu(\mathbb{Z}_{\oplus}) = 1/(1-r)$ and that $\mu(\{0\}) = 1$.

Example 5. A Binomial Measure. Let n be a positive integer and let $0 < p < 1$. Define μ as follows:

$$\mu(\mathcal{A}) = \sum_{k \in \mathcal{A} \cap \{0,1,\ldots,n\}} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k},$$

where ! denotes the factorial function, $j! = j(j-1) \cdots 1$.

Example 6. Bivariate Gaussian. Define a measure on $\mathbb{R}^2$ by

$$\mu(\mathcal{A}) = \int_{\mathcal{A}} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy.$$

Example 7. Uniform on a Ball in $\mathbb{R}^3$. Let $\mathcal{B}$ be the set of points in $\mathbb{R}^3$ that are within a distance 1 from the origin; this is called the unit ball in $\mathbb{R}^3$. (It looks like a ball centered on $(0,0,0)$.) Define a measure on $\mathbb{R}^3$ as follows:

$$\mu(\mathcal{A}) = \frac{3}{4\pi}\,\mu_{\text{Leb}}(\mathcal{A}\cap\mathcal{B}).$$

This measure yields $\mu(\mathbb{R}^3) = \mu(\mathcal{B}) = 1$. It looks like Lebesgue measure inside the unit ball but assigns measure 0 to any part of a set outside the ball.

Example 8. Not a measure. Suppose we define a set function that maps a set $\mathcal{A}$ to the smallest number (including ∞) that is $\geq$ every element of $\mathcal{A}$.[10] This arguably gives one way to quantify the size of the set but it is not a measure.

10 This is called the least upper bound, or supremum, of $\mathcal{A}$ and is denoted $\sup\mathcal{A}$. If $\mathcal{A}$ is finite, then $\sup\mathcal{A} = \max \mathcal{A}$, but some infinite sets like $]0,1[$ have no maximum. What is $\sup\,[0,1]$? $\sup\,]0,1[$? $\sup\mathbb{R}$?

The definition of a measure implies two useful properties that it is worthwhile to see. Suppose that μ is a measure. Then the following hold:

1. Monotonicity. If $\mathcal{A}\subset\mathcal{B}$, then $\mu(\mathcal{A}) \leq \mu(\mathcal{B})$.
2. Subadditivity. If $\mathcal{A}_1, \mathcal{A}_2, \ldots$ is a countable collection of sets – not necessarily disjoint – then

$$\mu\left(\bigcup_i \mathcal{A}_i\right) \leq \sum_i \mu(\mathcal{A}_i).$$

The monotonicity property was discussed in Section 1.1. It follows from additivity because $\mathcal{B} = \mathcal{A}\cup\operatorname{compl}(\mathcal{A})$ and the two pieces are disjoint. The subadditivity property can be understood intuitively: if we divide up a set into *overlapping* pieces, then the sum of the measure of the pieces must be greater than if we made the pieces disjoint since any overlap is counted more than once.

1.3. A Brief Digression on Infinitesimals

Although the calculus was developed independently by Newton and Leibniz, much of the notation we use today comes from Leibniz. Of particular relevance for our purposes are the "infinitesimals" like dx and dy that appear in the expressions for derivatives and integrals: df/dx and $\int f(x)\,dx$. They act as semantic placeholders that tell us which variables we are operating with respect to, but they also have a deeper meaning that can be quite enlightening.

Let's start with $f(x)\,dx$. The definition of the integral that you used in calculus approximates the area under the curve $f(x)$ by successively smaller rectangles. The expresion $f(x)\,dx$ is the area of a rectangle with height $f(x)$ and "infinitesimal" width dx that extends from x to $x + dx$. Thus, $f(x)\,dx$ is an infinitesimal area under the curve which we sum up during the integration. In a multivariate integral, the term $f(x, y)\,dx\,dy$ is the infinitesimal volume of a rectangular solid of height $f(x, y)$, horizontal width dx from x to $x + dx$, and vertical width dy from y to $y + dy$. The integral again adds up these infinitesimal values, yielding the volume under the two dimensional surface $f(x, y)$.

An infinitesimal quantity is one that is not zero but is arbitrarily – infinitely – small. From a rigorous mathematical viewpoint, this is nonsensical. A real number is either small but non-zero or is zero; there is nothing in between. What is actually meant is a more complicated limiting process that defines the shrinking of the rectangles. However, the idea of the infinitesimal is conceptually powerful. It explains an integral as a sum of quantities in the right units (areas for a one-dimensional integral, volumes for a two-dimensional integral, and so forth). It provides mnemonics for changes of variables and other transformations. It allows us to treat curves that change on arbitrarily small scales as a sequence of instantaneous values.

So what does something like dx mean? It is the length of an infinitesimal interval starting at x. When we see dx, it tells us three things: which variable we are dealing with, what value of that variable we are considering, and the length of an infinitesimal at that value. When we integrate with respect to measures in the subsections below, you will see an expression of the form $\mu(dx)$ where μ is a measure and dx is an infinitesimal. This is a short-hand notation for $\mu\Big([x, x + dx[\Big)$, that is, the measure of an infinitesimal interval starting at x. While more precise, the longer version is rather cumbersome, especially when repeated many many times. The beauty of infinitesimals is that this shorthand holds together nicely. For example, for Lebesgue measure, $\mu_{\text{Leb}}\Big([x, x + dx[\Big) = dx$, so we write $Lebesgue(dx) = dx$, and the integral with respect to Lebesgue measure, defined below, reduces to the ordinary integral.[11] In two and three dimensions, we use expressions of the form $\mu(dx\,dy)$ and $\mu(dx\,dy\,dz)$, which mean respectively $\mu\Big([x, x + dx[\ \times\ [y, y + dy[\Big)$ and $\mu\Big([x, x + dx[\times [y, y + dy[\times [z, z + dz[\Big)$. You can see why we prefer the shorthand. But again things work out because for Lebesgue measure

11 Actually, the integral with respect to Lebesgue measure is somewhat more general in that one can integrate a wider class of functions.

on $\mathbb{R}^2$ and $\mathbb{R}^3$ we have $\mu_{\text{Leb}}(dx\,dy) \equiv \mu_{\text{Leb}}\Big([x, x+dx[\ \times\ [y, y+dy[\Big) = dx\,dy$ and $\mu_{\text{Leb}}(dx\,dy\,dz) \equiv \mu_{\text{Leb}}\Big([x, x+dx[\times[y, y+dy[\times[z, z+dz[\Big) = dx\,dy\,dz$.

12 I say signed area because, as you may recall, when f dips below the x-axis the corresponding area counts negatively towards the integral.

1.4. Integration with respect to a Measure I: The Idea

In calculus, you learned that for a real-valued function f defined on a subset of $\mathbb{R}$, the *integral* of f, $\int f(x)\,dx$, is the (signed) [12] area between the curve $f(x)$ and the x-axis. Using measures, we can define a more powerful integral that reduces exactly to the Reimann integral where appropriate. Here, we give a conceptual overview of the integral with examples that show you what it means and how to compute it. In section 1.5, we describe the properties of the integral These two subsections should be sufficient for a strong working understanding of how to integrate. Section 1.6 gives the details of how this integral is constructed, which can help one understand exactly what is being computed.

We will consider functions $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is any set and a measure μ on $\mathcal{X}$ and compute the integral of f with respect to μ, denoted $\int f(x)\,\mu(dx)$. This is a real number associated with the function. The "x" in $\int f(x)\,\mu(dx)$ is just a dummy variable and can be replaced by anything else. One way to think about the integral is as a sum of infinitesimal areas $f(x)\,\mu(dx)$ over all x. Here, $f(x)$ is a height and $\mu(dx)$ is a corresponding width; the measure enters only in the latter.

We begin with four examples which show that there is little new here.

1. For any function f,

$$\int g(x)\,\mu_{\text{Leb}}(dx) = \int g(x)\,dx.$$

   In other words, integrals with respect to Lebesgue measure are computed exactly the same way that you learned in calculus. In terms of infinitesimals, $\mu_{\text{Leb}}(dx) = dx$, so the "sum of infinitesimals" we get is the same as you are used to.

2. For any function f,

$$\int g(x)\,\delta_{\{\alpha\}}(dx) = g(\alpha).$$

Consider the infinitesimal $\delta_{\{\alpha\}}(dx)$ as x ranges over $\mathbb{R}$. If $x \neq \alpha$, then the infinitesimal interval $[x, x+dx[$ does not contain α, so

$$\delta_{\{\alpha\}}(dx) \equiv \delta_{\{\alpha\}}\Big([x, x+dx[\Big) = 0.$$

If $x = \alpha$, $\delta_{\{\alpha\}}(dx) \equiv \delta_{\{\alpha\}}\Big([x, x+dx[\Big) = 1$. Thus, when we add up all of the infinitesimals, we get $g(\alpha) \cdot 1$.

3. For any function f,

$$\int g(x)\, \delta_{\mathbb{Z}}(dx) = \sum_{i \in \mathbb{Z}} g(i).$$

Again, consider the infinitesimal $\delta_{\mathbb{Z}}(dx)$ as x ranges over $\mathbb{R}$. If $x \notin \mathbb{Z}$, then the infinitesimal interval $[x, x+dx)$ does not intersect $\mathbb{Z}$ so $\delta_{\mathbb{Z}}(dx) \equiv \delta_{\mathbb{Z}}\Big([x, x+dx[\Big) = 0$. If $x \in \mathbb{Z}$, $\delta_{\mathbb{Z}}(dx) \equiv \delta_{\mathbb{Z}}\Big([x, x+dx[\Big) = 1$. (It is not more than one because an infinitesimal interval can contain at most one integer.) The term $g(x)\, \delta_{\mathbb{Z}}(dx)$ is thus $g(x)$ if $x \in \mathbb{Z}$ and 0 otherwise. When we add up all of the infinitesimals over x, we get the sum above.

4. Suppose $\mathcal{C}$ is a countable set. We can define counting measure on $\mathcal{C}$ to map $\mathcal{A}$ to $\#\mathcal{A} \cap \mathcal{C}$. For any function f,

$$\int g(x)\, \delta_{\mathcal{C}}(dx) = \sum_{v \in \mathcal{C}} g(v),$$

using the same basic argument as in the above example.

What we have just learned is that integrals with respect to Lebesgue measure are just ordinary integrals and that integrals with respect to Counting measure is just ordinary summation. So, in essence, we have a syntactic device for unifying sums and integrals to the same notation.

The next step is to consider measures built from Lebesgue and Counting measure.

1. Suppose μ is a measure that satisfies $\mu(dx) = f(x)\, \mu_{\mathsf{Leb}}(dx)$, then for any function g,

$$\int g(x)\, \mu(dx) = \int g(x)\, f(x)\, \mu_{\mathsf{Leb}}(dx) = \int g(x)\, f(x)\, dx.$$

As described in section 3, we say that f is the density of μ with respect to Lebesgue measure in this case.

2. Suppose μ is a measure that satisfies $\mu(dx) = p(x)\,\delta_{\mathcal{C}}(dx)$ for a countable set $\mathcal{C}$, then for any function g,

$$\int g(x)\,\mu(dx) = \int g(x)\,p(x)\,\delta_{\mathbb{C}}(dx) = \sum_{v \in \mathcal{C}} g(v)\,f(v).$$

   Here, we say that p is the density of μ with respect to Counting measure on $\mathcal{C}$.

Except for #, all of the measures we will use in this course will have a density with respect to Lebesgue measure or with respect to Counting measure on some set. Therefore, integration with respect to measures is nothing new except for a nicely unified way to express summation.

1.5. Properties of the Integral

A function is said to be *integrable* with respect to μ if $\int |f(x)|\,\mu(dx) < \infty$. An integrable function has a well-defined and finite integral. If $f \geq 0$, the integral is always well-defined but may be ∞. (Consider 1 integrated with respect to Lebesgue measure on $\mathbb{R}$.)

As in calculus, we sometimes wish to integrate over a subset of the domain. Suppose μ is a measure on $\mathcal{X}$, $\mathcal{A} \subset \mathcal{X}$, and g is a real-valued function on $\mathcal{X}$. We define the integral of g over the set $\mathcal{A}$, denoted by $\int_{\mathcal{A}} g(x)\,\mu(dx)$, to be simply the integral of g using only those x that are elements of $\mathcal{A}$. This is given by

$$\int_{\mathcal{A}} g(x)\,\mu(dx) = \int g(x)\,1_{\mathcal{A}}(x)\,\mu(dx).$$

Notice that the points outside of $\mathcal{A}$ contribute nothing to the integral. Because $1_{\mathcal{X}} = 1$, we have that $\int_{\mathcal{X}} f(x)\,\mu(dx) = \int f(x)\,\mu(dx)$. Again, this is syntactic only; there is nothing new here.

For a practical working mastery of the integral, what you really need to know is how to manipulate it. Integrals of any kind are just sums, and they satisfy all the properties that a sum does. For example, $|2-3+4-5| \leq |2|+|-3|+|4|+|-5|$ because the absolute values in the latter prevent the negatives from reducing the total sum. This same idea translates to the property $|\int f(x)\,\mu(dx)| \leq \int |f(x)|\,\mu(dx)$; again, the absolute values inside the integral prevent any cancelation that would reduce the magnitude of the result. Similarly, $1+4+7+9 \leq 2+6+8+12$ because each term on the left is $\leq$ the corresponding term on the right. This translates to the property $f \le g$ implies $\int f(x)\,\mu(dx) \le \int g(x)\,\mu(dx)$.

When the integral is defined as in the next subsection, the following properties hold for every measure μ. In the list below, μ is a measure on $\mathcal{X}$, $\mathcal{A}, \mathcal{B} \subset \mathcal{X}$, $c \in \mathbb{R}$ and f and g are integrable functions.

1. *Constant Functions.* $\int_{\mathcal{A}} c\,\mu(dx) = c \cdot \mu(\mathcal{A})$.
2. *Linearity.*

$$\int_{\mathcal{A}} cf(x)\,\mu(dx) = c \int_{\mathcal{A}} f(x)\,\mu(dx)$$
$$\int_{\mathcal{A}} (f(x) + g(x))\,\mu(dx) = \int_{\mathcal{A}} f(x)\,\mu(dx) + \int_{\mathcal{A}} g(x)\,\mu(dx).$$

3. *Monotonicity.* If $f \le g$, then $\int_{\mathcal{A}} f(x)\,\mu(dx) \le \int_{\mathcal{A}} g(x)\,\mu(dx)$ for every set $\mathcal{A}$. This implies the following common special cases.
   - If $f \ge 0$, $\int f(x)\,\mu(dx) \ge 0$.
   - If $f \ge 0$ and $\mathcal{A} \subset \mathcal{B}$, $\int_{\mathcal{A}} f(x)\,\mu(dx) \le \int_{\mathcal{B}} f(x)\,\mu(dx)$.
4. *Null Sets.* If $\mu(\mathcal{A}) = 0$, then $\int_{\mathcal{A}} f(x)\,\mu(dx) = 0$.
5. *Absolute Values.* $|\int f(x)\,\mu(dx)| \le \int |f(x)|\,\mu(dx)$.
6. *Monotone Convergence.* If $0 \le f_1 \le f_2 \le \cdots$ is an increasing sequence of integrable functions that converge to f, then

$$\lim_{k\to\infty} \int f_k(x)\,\mu(dx) = \int f(x)\,\mu(dx).$$

7. *Linearity in region of integration.* If $\mathcal{A} \cap \mathcal{B} = \emptyset$,

$$\int_{\mathcal{A}\cup\mathcal{B}} f(x)\,\mu(dx) = \int_{\mathcal{A}} f(x)\,\mu(dx) + \int_{\mathcal{B}} f(x)\,\mu(dx).$$

To keep all of the above properties straight in your mind, it may help to think about the signed area under a curve. Each property translates into a physical statement about areas. For example, Monotonicity says that if one area contains another, it is bigger. The Absolute Values property results from the fact that any negative values subtract from the total area; the absolute values eliminate this. The Monotone Convergence property says that if we fill up the area under a curve f, building up in a series of steps, then this area gets closer and closer to the total area under f.

1.6. Integration with respect to a Measure II: The Details

To define an integral means to create a consistent and specific algorithm for computing the integral of any appropriate function.[13] As you know from experience, some functions are easier to integrate than others. For example, to integrate a multiple of the indicator function $1_{[0,1]}$ requires nothing special—we simply multiply height by width to get the area under the curve. For functions like this, we can define the integral to our complete satisfaction, so we start out defining the integral of simple functions like this. The next trick is to extend the definition of the integral to more complicated functions in a way that the integral behaves like we expect it to. Below, we define the integral with respect to a measure in three steps.

The integral introduced in calculus is called the Reimann integral. It is defined by approximating the area under the curve $f(x)$ with smaller and smaller rectangles. To define $\int f(x)\,\mu(x)$, we will do almost the same thing, but our method for approximating a function f is a bit different than filling in under the curve with rectangles. [14] On your first reading, you may want to skip this subsection. Come back to it later for deeper insights when you feel more comfortable with the integrals.

Step 1. Define the integral for simple functions.

A *simple function*, is any function that takes only a finite number of different values. All constant functions are simple functions because they take only one value. The indicator function of a set $\mathcal{A} \subset \mathcal{X}$ is a simple function because $1_{\mathcal{A}}$ takes only two values: 0 and 1. Any constant c times an indicator function, that is $c \cdot 1_{\mathcal{A}}$, is also a simple function because it takes at most the values 0 and c. Similarly, given disjoint sets $\mathcal{A}_1$ and $\mathcal{A}_2$, the linear combination $c_1 \cdot 1_{\mathcal{A}_1} + c_2 \cdot 1_{\mathcal{A}_2}$ is a simple function which can take the values 0, c_1, c_2.[15] In fact, any simple function f can be expressed as a linear combination of a finite number of indicator functions. That is, if f is *any* simple function on $\mathcal{X}$, then there is some finite integer n, there are non-zero constants $c_1, \ldots, c_n$, and there are *disjoint* sets $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{X}$ such that

$$f = c_1 \cdot 1_{\mathcal{A}_1} + c_2 \cdot 1_{\mathcal{A}_2} + \cdots + c_n \cdot 1_{\mathcal{A}_n}.$$

Because the sets are disjoint, this function takes the $n+1$ values $0, c_1, \ldots, c_n$.

13 For mathematicians to be satisfied, the algorithm need only work in principle; we may not be able to implement it in practice.

14 The construction of the Reimann integral is usually rushed in calculus courses, and it is hardly the most exciting part of the material in any case. If you do not remember how it goes, do not worry. The method we will use is actually much simpler.

15 If $\mathcal{A}_1$ and $\mathcal{A}_2$ were not disjoint, we could define $\mathcal{B}_1 = \mathcal{A}_1 - \mathcal{A}_2$, $\mathcal{B}_2 = \mathcal{A}_2 - \mathcal{A}_1$, and $\mathcal{B}_3 = \mathcal{A}_1 \cap \mathcal{A}_2$ and then the function is equal to $c_1 \cdot 1_{\mathcal{B}_1} + c_2 \cdot 1_{\mathcal{B}_2} + (c_1 + c_2) \cdot 1_{\mathcal{B}_3}$. Make sure you understand why this is true. (Draw a picture in one example.)

16 ATTN Metapost figure with simple function and areas illustrated.

17 Remember: The variable of integration is a dummy variable.

So, if $f\colon \mathcal{X} \to \mathbb{R}$ is a simple function as just defined, what is the integral of f? Keeping in mind that the integral represents the (signed) area under the curve, all we have to do for each segment is to multiply the height c by the width $\mu(\mathcal{A})$ and add them all up. Specifically, if

$$f = c_1 \cdot 1_{\mathcal{A}_1} + c_2 \cdot 1_{\mathcal{A}_2} + \cdots + c_n \cdot 1_{\mathcal{A}_n}. \tag{M.1}$$

then

$$\int f\,\mu(dx) = c_1 \cdot \mu(\mathcal{A}_1) + c_2 \cdot \mu(\mathcal{A}_2) + \cdots + c_n \cdot \mu(1_{\mathcal{A}_n}). \tag{M.2}$$

This integral can be a real number, ∞, or $-\infty$ depending on the sets and the constants defining f. Take a look at the figure 16 and the following examples to get a feel for the argument,

Example 1. Suppose that μ is Lebesgue measure and that f is defined as follows:

$$f(x) = \begin{cases} 1 & \text{if } -1 < x < 1 \\ 3 & \text{if } 1 \le x < 4 \\ 5 & \text{if } 4 \le x < 10 \\ 2 & \text{if } -4 < x \le -1 \\ 4 & \text{if } -10 < x \le -4 \\ 0 & \text{otherwise.} \end{cases}$$

Find $\int f(t)\,\mu(dt)$.[17]

The definition of f above makes clear that it is a simple function because it takes only 6 possible values. Let's express it in the form of equation (M.1):

$$f = 1_{]-1,1[} + 3 \cdot 1_{[1,4[} + 5 \cdot 1_{[4,10[} + 2 \cdot 1_{]-4,-1]} + 4 \cdot 1_{]-10,4]}. \tag{M.3}$$

Because μ is Lebesgue measure on $\mathbb{R}$, $\mu(dx) = dx$; so, $\int f\,\mu(dx) = \int f\,dx$, the familiar integral from calculus. The area under f can now be computed (draw a picture) by calculating the area under each separate piece and adding all the areas together. As above, we have

$$\begin{aligned} \int f\,\mu(dx) &= \mu\big(]-1,1[\big) + 3 \cdot \mu\big([1,4[\big) + 5 \cdot \mu\big([4,10[\,\big) + \\ &\qquad 2 \cdot \mu\big(]-4,-1]\big) + 4 \cdot \mu\big(]-10,4]\big) \\ &= 2 + 3 \cdot 3 + 5 \cdot 6 + 2 \cdot 3 + 4 \cdot 6 \\ &= 71. \end{aligned}$$

Example 2. Suppose f is defined as above and that $\mu = \delta_{\mathbb{Z}}$, counting measure on $\mathbb{Z}$. Find $\int f(x)\,\mu(dx)$.

The representation of f in terms of indicators is given in equation (M.3), and we need only compute the measures of the sets as in equation (M.2); see also the first line of the previous display. Since $\mu(\mathcal{A})$ is the number of integers in $\mathcal{A}$, we have that $\mu\Big(]-1,1[\Big) = 1$, $\mu\Big([1,4[\Big) = 3$, $\mu\Big([4,10[\Big) = 6$, and so forth. Hence, $\int f\,\mu(dx) = 1\cdot 1 + 3\cdot 3 + 5\cdot 6 + 2\cdot 3 + 4\cdot 6 = 70$.

Example 3. Suppose that f is any simple function and that $\mu = \delta_{\{0\}}$, a point mass at 0. Find $\int f(y)\,\mu(dy)$.

Here, μ of any set is 0 unless that set contains 0. So in the representation of f of equation (M.1), only one of the sets $\mathcal{A}_i$ will contain 0. The measure of this set will be 1, and the constant multiplying the corresponding indicator is just $f(0)$. It follows that $\int f(y)\,\delta_{\{0\}}(dy) = f(0)$ for any f.

Step 2. Define the integral for general non-negative functions, approximating the general function by simple functions.

Having learned how to compute the integral of any simple function, we turn next to a general non-negative function $f\colon \mathcal{X} \to [0,\infty[$. The idea is that we can approximate[18] such an f arbitrarily well by some non-negative simple function that is $\leq f$, as illustrated in the marginal figure. The argument[19] that this claim is true is enlightening, but you can skip it without worry on the first reading.

If f and g are non-negative functions on the same domain then the integrals of f and g include no "negative areas"; hence, if $g \leq f$, we should have $\int g(x)\,\mu(dx) \leq \int f(x)\,\mu(dx)$. Indeed, this is true as can be seen from the definition above if f and g are both simple functions. We define the integral to make this true statement true. Thus, if $f\colon \mathcal{X} \to [0,\infty[$ is a general functions and $0 \leq s \leq f$ is a simple function, we require that $\int s(x)\,\mu(dx) \leq \int f(x)\,\mu(dx)$. The closer that s approximates f, the closer we expect $\int s(x)\,\mu(dx)$ and $\int f(x)\,\mu(dx)$ to be.

This is how we compute the integral of f. Loosely speaking, we define $\int f(x)\,\mu(dx)$ to be the largest value $\int s(x)\,\mu(dx)$ for a simple function $0 \leq s \leq f$. To be more precise, we define the integral $\int f(x)\,\mu(dx)$ to be the smallest value I such that $\int s(x)\,\mu(dx) \leq I$ for all simple functions

18 ATTN: Metapost figure of successive approximations.

19 To see this, choose some large integers k, and define $\mathcal{A}_i = f^{-1}\Big([i/2^k, (i+1[\ /2^k)\Big)$ for $0 \leq i < k2^k$ and $\mathcal{A}_{k2^k} = f^{-1}\Big([k,\infty[\Big)$, where as usual $f^{-1}(\mathcal{B})$ refers to the inverse image of the set $\mathcal{B}$. Define a simple function s_k by

$$s_k(x) = \sum_{i=0}^{k2^k} \left(\frac{i}{2^k}\right) 1_{\mathcal{A}_i}.$$

Since $s(x) = i/2^k \leq f(x)$ for all $x \in \mathcal{A}_i$, we have $s \leq f$. Also, $|s_k(x) - f(x)| \leq 2^{-k}$ except possibly when $x \in \mathcal{A}_{k2^k}$. and as k gets larger and larger, each x is eventually outside $\mathcal{A}_{k2^k}$.

$0 \leq s \leq f$.[20]

20 This is the least upper bound (supremum) as defined in note 10. In other words, $\int f(x)\,\mu(dx)$ is the least upper bound of the set of numbers $\int s(x)\,\mu(dx)$ such that s is simple and $0 \leq s \leq f$. We need the least upper bound because there might not be a largest $\int s(x)\,\mu(dx)$.

Step 3. Define the integral for general real-valued functions by separately integrating the positive and negative parts of the function.

Finally, if $f\colon \mathcal{X} \to \mathbb{R}$ is a general function, we can define its *positive part* f^+ and its *negative part* f^- by[21]

21 ATTN: Metapost Figure. Function with positive and negative parts

$$f^+(x) = \max(f(x), 0)$$
$$f^-(x) = \max(-f(x), 0).$$

Notice that both f^+ and f^- are non-negative functions (so we know how to integrate them) and that $f = f^+ - f^-$. Consequently, we define

$$\int f(x)\,\mu(dx) = \int f^+(x)\,\mu(dx) - \int f^-(x)\,\mu(dx).$$

This is a well-defined number (possibly infinite) if and only if at least one of f^+ and f^- has a finite integral.

This defines the integral of a function f with respect to a measure μ. The advantage of seeing the construction is that we understand exactly what the integral means, but we can operate just fine using only the examples and properties described in the previous two subsections. Fortunately, one can show that this construction implies that all of those properties are true.

2. Some Important Examples

We saw a variety of examples of measures above. However, every measure that we will encounter in this course will be built from one of the three most important examples, which we summarize in the following:

1. Cardinality. This is denoted by $\#(\mathcal{A})$ for any set $\mathcal{A}$ and can be defined as a measure on any domain.
2. Counting measure.

   Suppose $\mathcal{C}$ is a countable set. Then, the counting measure with respect to $\mathcal{C}$, $\delta_{\mathcal{C}}$, is defined by $\delta_{\mathcal{C}}(\mathcal{A}) = \#\mathcal{A} \cap \mathcal{C}$. Counting measure on the integers, $\delta_{\mathbb{Z}}$, and the point mass at a point x, $\delta_{\{x\}}$, are special cases.

   Integrals with respect to counting measure are just sums. That is,

$$\int f(x)\,\delta_{\mathcal{C}}(dx) = \sum_{z \in \mathcal{C}} f(z).$$

Or, put another way, sums and ordinary integrals are both just integrals, but with respect to different measures.

3. Lebesgue measure. This can be defined on any $\mathbb{R}^k$. Lebesgue measure on $\mathbb{R}$ gives length, Lebesgue measure on $\mathbb{R}^2$ gives area, and Lebesgue measure on $\mathbb{R}^3$ gives volume, all three very familiar. We always denote Lebesgue measure by μ_{Leb} and let context indicate the appropriate domain, but it will be rare that we have to refer to it explicitly.

   Lebesgue measure on any $\mathbb{R}^k$ satisfies the following properties:

   A. A hyper-rectangle[22] $]a_1, b_1[\, \times \cdots \times \,]a_k, b_k[$ in $\mathbb{R}^k$ has Lebesgue measure $(b_1 - a_1) \cdots (b_k - a_k)$.
   B. If $\mathcal{C}$ is any countable set, $\mu_{\mathsf{Leb}}(\mathcal{C}) = 0$.
   C. If $\mathcal{B}$ is the result of shifting the set $\mathcal{A}$ by some amount (e.g., $[4, 5]$ is obtained by shifting $[0, 1]$ by 4 in $\mathbb{R}$), then $\mu_{\mathsf{Leb}}(\mathcal{B}) = \mu_{\mathsf{Leb}}(\mathcal{A})$.

   Integrals with respect to Lebesge measure are just the ordinary integrals you are familiar with from calculus. In particular, $\mu_{\mathsf{Leb}}(dx) = dx$ on $\mathbb{R}$, $\mu_{\mathsf{Leb}}(dx\,dy) = dx\,dy$ on $\mathbb{R}^2$, $\mu_{\mathsf{Leb}}(dx\,dy\,dz) = dx\,dy\,dz$ on $\mathbb{R}^3$, and $\mu_{\mathsf{Leb}}(dx_1 \cdots dx_k) = dx_1 \cdots dx_k$ on $\mathbb{R}^k$.

22 A hyper-rectangle in $\mathbb{R}$ is just an interval; in $\mathbb{R}^2$ it is a rectangle; and in $\mathbb{R}^3$ it is a rectangular solid.

3. How to make new measures from old ones

Sums and Multiples. Consider the point mass measures at 0 and 1, $\delta_{\{0\}}$ and $\delta_{\{1\}}$, and construct a two new measures on $\mathbb{R}$ $\mu = \delta_{\{0\}} + \delta_{\{1\}}$ defined by

$$\mu(\mathcal{A}) = \delta_{\{0\}}(\mathcal{A}) + \delta_{\{1\}}(\mathcal{A}),$$

and $\nu = 4\delta_{\{0\}}$ defined by

$$\nu(\mathcal{A}) = 4 \cdot \delta_{\{0\}}(\mathcal{A}),$$

for all $\mathcal{A} \subset \mathcal{X}$. We say that μ is the sum of $\delta_{\{0\}}$ and $\delta_{\{1\}}$ and that ν is a multiple of $\delta_{\{0\}}$ by a factor or 4. The measure μ counts how many elements of $\{0, 1\}$ are in its argument. (We will denote counting measures like this by a common convention; specifically, $\mu = \delta_{\{0,1\}}$ in this case.) Notice that $\delta_{\mathbb{Z}}$, the counting measure of the integers can be re-expressed as a sum of all pointmasses $\delta_{\{i\}}$ for integers i:

$$\delta_{\mathbb{Z}} = \sum_{i=-\infty}^{\infty} \delta_{\{i\}}.$$

Similarly, we can also write ν as a sum

$$\nu = \delta_{\{0\}} + \delta_{\{0\}} + \delta_{\{0\}} + \delta_{\{0\}}.$$

Because we have defined measures to be non-negative, a multiple of a measure is only a measure if the constant factor is non-negative.

By combining the operations of summation and multiplication, we can construct measures like $4\delta_{\{0\}} + 3\delta_{\{1\}}$ that are linear combinations of other measures (with positive coefficients). For example, the geometric measure in example ATTN above can be written as $\sum_{i=0}^{\infty} r^i \, \delta_{\{i\}}$.

Restriction to a Subset. Consider Lebesgue measure on $\mathbb{R}$, μ_{Leb}. We can define a new measure on $]0, 1[$ which maps a set $\mathcal{A} \subset \,]0, 1[$ to $\mu_{\mathsf{Leb}}(\mathcal{A})$. This is effectively the same measure but it is restricted to subsets of $]0, 1[$ only. In general, suppose μ is a measure on $\mathcal{X}$ and $\mathcal{B} \subset \mathcal{X}$. We can define a new measure on $\mathcal{B}$ which maps $\mathcal{A} \subset \mathcal{B}$ to $\mu(\mathcal{A})$. This is called the restriction of μ to the set $\mathcal{B}$.

The Measure Induced by a Function. Consider Lebesgue measure on $\mathbb{R}$ and the function $g\colon \mathbb{R} to \,]0, \infty[$ defined by $g(x) = 10^x$. The function g connects subsets of $(0, \infty)$ to subsets of $\mathbb{R}$ via the inverse image. For example, the inverse image of $]1, 100[\, \subset \,]0, \infty[$ is $]0, 2[\, \subset \mathbb{R}$ because these are the exponents of 10 that produce the values 1 through 100. We can use this relationship to construct a new measure on $(0, \infty)$: $\nu(\mathcal{A}) = \mu_{\mathsf{Leb}}(g^{-1}(\mathcal{A}))$ for $\mathcal{A} \subset \,]0, \infty[$.

Thought Question Find $\nu\big(]1, 2[\big)$, $\nu\big(]1, 100[\big)$, $\nu\big(]1, 10000[\big)$, and $\nu\big(]100, 10000[\big)$.

In general, suppose μ is a measure on $\mathcal{X}$ and $g\colon \mathcal{X} \to \mathcal{Y}$ is a function. We can use μ and g to define a new measure ν on $\mathcal{Y}$ by

$$\nu(\mathcal{A}) = \mu(g^{-1}(\mathcal{A}))$$

for $\mathcal{A} \subset \mathcal{Y}$. We call this the measure *induced from μ by the function g*. Notice that we can think of the inverse image g^{-1} as mapping subsets of $\mathcal{Y}$ to subsets of $\mathcal{X}$, that is as a function $2^{\mathcal{Y}} \to 2^{\mathcal{X}}$. In that sense, the measure induced from μ by g can be written as $\mu \circ g^{-1}$. This technique allows us to carry over a measure defined on one space to another space with no extra work. It is very important in probability theory as we will see in Chapter 3.

Thought Question Suppose $h(x) = x^2$. Describe $\mu_{\text{Leb}} \circ h^{-1}$.

If $\nu = \mu \circ g^{-1}$ for some function, g, then we can make a useful substitution in integrals that illuminates the relationship between μ and ν. Specifically, $f(y)\,\nu(dy) = f(g(x))\,\mu(dx)$, for $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, so wherever we see one, we can substitute the other. In other words, for any function $f\colon \mathcal{Y} \to \mathbb{R}$,

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

Integrating a Density. Suppose μ is a measure on $\mathcal{X}$ and $f\colon \mathcal{X} \to \mathbb{R}$. We can define a new measure ν on $\mathcal{X}$ as follows:

$$\nu(\mathcal{A}) = \int_{\mathcal{A}} f(x)\,\mu(dx). \tag{M.4}$$

(Recall the definition of the integral over a set above.) Because of the linearity and convergence properties of the integral, ν satisfies all the properties of a measure including countable additivity.

We say that f is the *density* of the measure ν with respect to the measure μ. The name "density" seems an odd choice, but it does have a strong physical intuition which we will explore fully in Chapter 3.

We often turn this phrase around. If ν and μ are two measures for which the equation (M.4) holds for every $\mathcal{A} \subset \mathcal{X}$, we say that ν has a density f with respect to μ. This implies two useful results:

- $\mu(\mathcal{A}) = 0$ implies $\nu(\mathcal{A}) = 0$.
- $\nu(dx) = f(x)\,\mu(dx)$. In other words, we can substitute for $\nu(dx)$ in integrals: for real-valued g on $\mathcal{X}$, $\int g(x)\,\nu(dx) = \int g(x)\,f(x)\,\mu(dx)$.

Densities play an important role in probability theory.

Example 1. Suppose, given μ, we define ν by $\nu(\mathcal{A}) = \mu(\mathcal{A} \cap \mathcal{B})$ for some fixed set $\mathcal{B} \subset \mathcal{X}$. Then, ν assigns 0 measure to anything outside of $\mathcal{B}$; for any subset of $\mathcal{B}$, μ and ν are the same. We can rewrite ν as follows

$$\nu(\mathcal{A}) = \int_{\mathcal{A}} 1_{\mathcal{B}}(x)\,\mu(dx).$$

Thus, $1_{\mathcal{B}}(x)$ is the density of ν with respect to μ. That the density is 0 outside $\mathcal{B}$ reflects the fact that ν assigns 0 measure to anything outside $\mathcal{B}$. That the density is 1 inside $\mathcal{B}$ reflects that fact that μ and ν are the same for any subset of $\mathcal{B}$. We could write this relationship in terms of infinitesimals as follows:

$$\nu(dx) = \begin{cases} \mu(dx) & \text{if } x \in \mathcal{B} \\ 0 & \text{if } x \notin \mathcal{B}. \end{cases}$$

Example 2. The geometric measure defined above ATTN(xref) has a density with respect to $\delta_{\mathbb{Z}}$, and the density is the function p given by

$$p(x) = \begin{cases} r^x & \text{if } x \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases}$$

Example 3. The point mass $\delta_{\{0\}}$ also has a density with respect to $\delta_{\mathbb{Z}}$. To see this, notice that

$$\int g(x)\, \delta_{\mathbb{Z}}(dx) = \sum_{i=-\infty}^{\infty} g(i).$$

If p is the function given by

$$p(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\int g(x)\, p(x)\, \delta_{\mathbb{Z}}(dx) = g(0).$$

By taking $g = 1_{\mathcal{A}}$ for a subset $\mathcal{A}$ of $\mathcal{X}$, we get

$$\int_{\mathcal{A}} p(x)\, \delta_{\mathbb{Z}}(dx) = \delta_{\{0\}}(\mathcal{A}).$$

Thus, p is the density of $\delta_{\{0\}}$ with respect to $\delta_{\mathbb{Z}}$.

Example 4. Define a measure μ_G on $\mathbb{R}$ by

$$\mu_G(\mathcal{A}) = \int_{\mathcal{A}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\, dx.$$

This is called standard Gaussian measure. It has a density $f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$ with respect to Lebesgue measure on $\mathbb{R}$. Notice that $\mu_G(\mathbb{R}) = 1$ because the density decays so rapidly for large values of $|x|$.

4. Other Types of Measures

Suppose that μ is a measure on $\mathcal{X}$.

If $\mu(\mathcal{X}) = \infty$, we say that μ is an infinite measure. If $\mu(\mathcal{X}) < \infty$, we say that μ is a finite measure. If $\mu(\mathcal{X}) = 1$, we say that μ is a probability measure. We will talk much more about probability measures throughout the course.

If there is a countable set $\mathcal{S}$ such that $\mu(\mathcal{X} - \mathcal{S}) = 0$, we say that μ is a *discrete* measure. Equivalently, μ has a density with respect to counting measure on $\mathcal{S}$. If μ has a density with respect to Lebesgue measure, we say that μ is a *continuous* measure. If μ is neither continuous nor discrete, we say that μ is a mixed measure.