

Probability Theory on Finite Spaces

Michael Betancourt

January 2023

Table of contents

1	Finite Spaces	1
1.1	Subsets	2
1.2	Subset Operations	4
2	Measure and Probability Over Elements	7
3	Measure and Probability Over Subsets	12
4	Acknowledgements	17
	License	18

Despite its infamous reputation the foundations of probability theory are quite straightforward. Much of the mathematical difficulty arises only when we implement probability theory on elaborate spaces like the real numbers, and most of the philosophical difficulty arises only when trying to assign an interpretation to the mathematics. In this chapter I try to avoid this baggage by introducing the basics of abstract probability theory on the simplest, nontrivial space: a collection of a finite number of elements.

1 Finite Spaces

A **finite space** is a set containing a finite number of distinct elements,

$$X = \{x_1, \dots, x_N\}.$$

The numerical index here allows us to differentiate the between the N individual elements but it does not necessarily imply that the elements have any natural ordering to them. For example the space

$$X = \{x_1, x_2, x_3, x_4, x_5\}$$

contains five, generally unordered elements (Figure 1). If the elements are endowed with a natural ordering, and the ordering is relevant, then we say that the space is an **ordered space**.

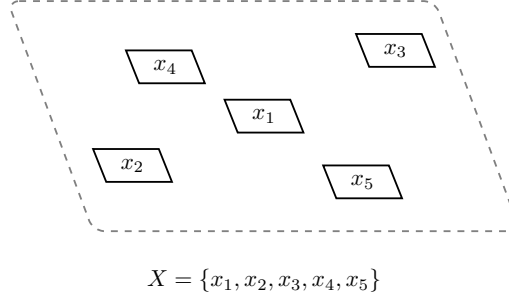


Figure 1: A finite space is a set containing of a finite number of elements.

In practical applications of probability theory the abstract elements x_n will model meaningful objects, but in this chapter I will avoid any particular interpretation and instead focus on the mathematical concepts. That said when X is intended to capture *all* of the objects of interest in a given application I will refer to it as the **ambient space**.

Once we've defined an ambient space we have various ways of organizing its individual elements and manipulating those organizations.

1.1 Subsets

A **subset** of X is any collection of elements in X . For example $x = \{x_1, x_3, x_4\}$ is a subset of $X = \{x_1, x_2, x_3, x_4, x_5\}$ (Figure 2). Importantly there is no notion of multiplicity in the concept of a subset, just membership: a subset can include an element x_n but it cannot include it multiple times.

If x is a subset of the ambient space X then we write $x \subset X$. When x might contain all of the elements of X , in which case $x = X$, then we write $x \subseteq X$.

Subsets are *recursive* objects. Selecting some elements from a subset yields a new subset. For example the collection $x' = \{x_3, x_4\}$ is both a subset of the previously introduced subset,

$$x' \subset \{x_1, x_3, x_4\}$$

, and the full ambient space, $x' \subset X$. More formally a subset x' if a subset of another subset x if all of the elements in x' are also in x .

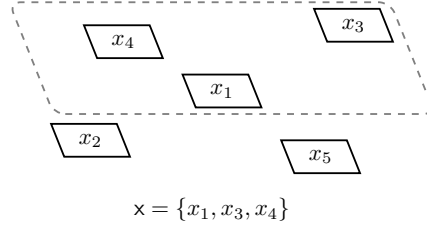


Figure 2: A subset $x \subset X$ is any collection of elements from the ambient space X .

Regardless of how many elements a finite space X contains three special subsets are always well-defined. The **empty set** $\emptyset = \{\}$ contains no elements at all. On the other hand the entire space itself can be considered a subset containing all of the elements. A subset containing a single element is denoted $\{x_n\}$ and referred to as an **atomic set**.

There are

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

ways to select n elements from a finite space of N total elements, and hence $\binom{N}{n}$ total subsets of size n . For example there is only one subset that contains no elements,

$$\binom{N}{0} = \frac{N!}{0!(N-0)!} = \frac{N!}{N!} = 1,$$

which is just the empty set. Similarly there is only one subset that contains all of the elements,

$$\binom{N}{N} = \frac{N!}{N!(N-N)!} = \frac{N!}{N!} = 1,$$

which is just the full space itself. On the other hand there are

$$\binom{N}{1} = \frac{N!}{1!(N-1)!} = N$$

distinct atomic sets that contain a single element, one for each element in X .

Counting all of the subsets of all possible sizes gives

$$\sum_{n=0}^N \binom{N}{n} = 2^N$$

possible subsets that we can construct from a finite space with N elements. In other words the collection of all subsets is itself a finite space with 2^N elements. We refer to this set as the **power set of X** and denote it 2^X .

When the ambient space, and all hence all of the possible elements that any subset might contain, is fixed the prefix “sub” is sometimes dropped, with all subsets simply referred to as sets. For example collections of elements might be denoted “sets” when one takes the ambient space for granted and “subsets” when one wants to explicitly acknowledge the context of the ambient space, but the precise terminology can vary strongly from context to context.

1.2 Subset Operations

We can always construct subsets element by element, but we can also construct them by manipulating existing subsets.

For example given a subset $x \subset X$ we can construct its **complement** by collecting all of the elements in X that are not already in x . The atomic set $x = \{x_3\}$ contains the lone element x_3 and its complement contains the remaining elements (Figure 3)

$$x^c = \{x_1, x_2, x_4, x_5\}.$$

Moreover the complement of the empty set is the entire space, $\emptyset^c = X$, and the complement of the full space is the empty set, $X^c = \emptyset$.

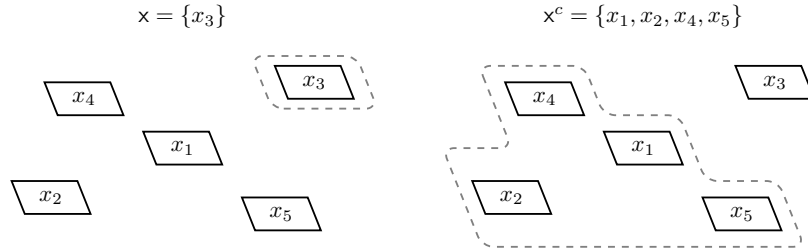


Figure 3: The complement of a subset x is the subset x^c consisting of all elements in the ambient space that are not in x .

More formally the construction of complementary subsets defines a *unary* function that takes in a subset as input and outputs the complementary subset,

$$\begin{aligned} \cdot^c : 2^X &\rightarrow 2^X \\ x &\mapsto x^c. \end{aligned}$$

The top line of this notation denotes the input space and output space of the operation, here both the power set, while the bottom line denotes the action on a particular input, here a

subset mapped into its complement. For example applying the complement operator to the subset $x = \{x_2, x_5\}$ gives

$$x^c = \{x_2, x_5\}^c = \{x_1, x_3, x_4\}.$$

We can also construct subsets from multiple subsets. Consider, for example, two subsets $x_1 = \{x_1, x_4\}$ and $x_2 = \{x_1, x_5\}$ (Figure 4). The collection of all elements that are contained in *either* subset is itself a subset, as is the collection of all elements that are contained only in *both* subsets. These are referred to as the **union**, $x_1 \cap x_2$, and **intersection**, $x_1 \cup x_2$, respectively (Figure 5). For example

$$\{x_1, x_3, x_5\} \cup \{x_2, x_3\} = \{x_1, x_2, x_3, x_5\}$$

and

$$\{x_1, x_3, x_5\} \cap \{x_2, x_3\} = \{x_3\}.$$

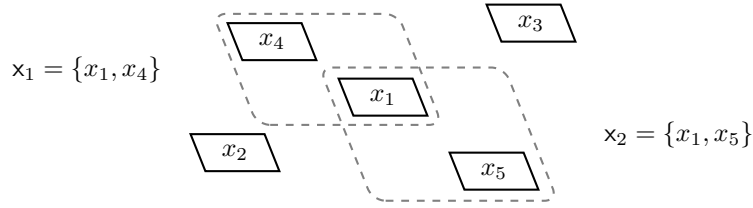


Figure 4: We can manipulate two subsets into a new subset in various ways.

Two subsets are **disjoint** if they don't share any elements; in this case their intersection is the empty set,

$$x_1 \cap x_2 = \emptyset.$$

The union and intersection of a subset with itself returns that subset,

$$x \cup x = x \cap x = x.$$

Because the empty set does not contain any elements its union with any subset returns back that subset,

$$x \cup \emptyset = \emptyset \cup x = x,$$

$$x_1 \cup x_2 = \{x_1, x_4, x_5\}$$

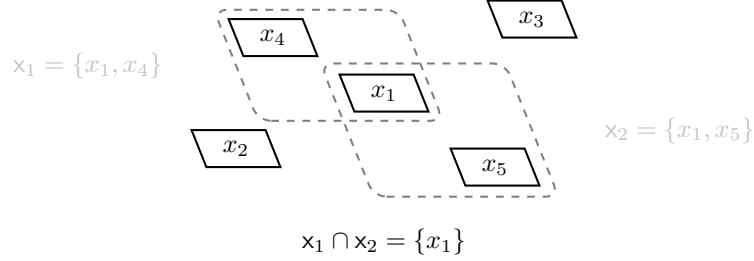


Figure 5: The union of two subsets, $x_1 \cup x_2$, is a subset containing all of the elements in either input subset. On the other hand the intersection of two subsets, $x_1 \cap x_2$, is a subset containing just the elements that occur in both input subsets.

and its intersection with any subset returns the empty set,

$$x \cap \emptyset = \emptyset \cap x = \emptyset.$$

Similarly the union of any subset with the full space returns the full space,

$$x \cup X = X \cup x = X,$$

and the intersection of any subset with the full space returns back the subset,

$$x \cap X = X \cap x = x.$$

Because they require two inputs the union and intersection define *binary* functions that consume *two* input subsets and return a single output subset,

$$\begin{aligned} \cdot \cup \cdot : 2^X \times 2^X &\rightarrow 2^X \\ x_1, x_2 &\mapsto x_1 \cup x_2 \end{aligned}$$

and

$$\begin{aligned} \cdot \cap \cdot : 2^X \times 2^X &\rightarrow 2^X \\ x_1, x_2 &\mapsto x_1 \cap x_2. \end{aligned}$$

Here $2^X \times 2^X$ denotes the space consisting of all *pairs* of subsets.

2 Measure and Probability Over Elements

Measure theory, and its special case of probability theory, is often burdened with intricate if not mysterious interpretations. From a mathematical perspective, however, measure theory simply concerns the consistent **allocation** of some abstract quantity across the ambient space.

Consider a *reservoir* of some real-valued, positive, and *conserved* quantity, $M \in \mathbb{R}^+$ (Figure 6). Because M is conserved any amount m_n that is allocated to the element $x_n \in X$ has to be *depleted* from the reservoir. That said in general the total content in the reservoir M can be infinite.

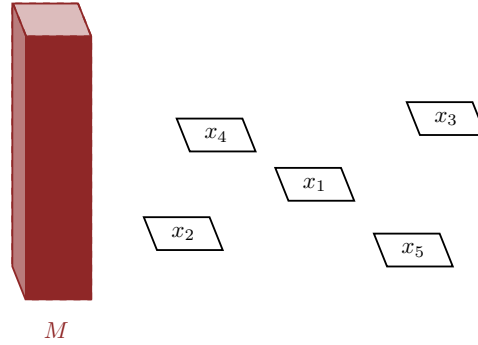


Figure 6: Measure theory concerns the allocation of some real-valued and positive quantity M over the individual elements of the ambient space.

To make the mathematics as useful as possible we will avoid endowing M with any particular meaning for the time being. Instead interpretations will arise only when we *apply* measure theory to model particular systems.

An *exhaustive* allocation of M across the ambient spaces ensures that the reservoir is completely emptied. In another words all of M has to be allocated to the elements $x_n \in X$.

For example consider the ambient space $X = \{x_1, x_2, x_3, x_4, x_5\}$. If we allocate m_1 to x_1 then that leaves $M - m_1$ remaining to allocate to the other four elements (Figure 7a). Allocating m_2 to x_2 depletes the reservoir a bit more (Figure 7b). At the end we have to allocate all

$$M - m_1 - m_2 - m_3 - m_4$$

that remains to x_5 (Figure 7e).

A **measure** is any consistent allocation of the quantity M to the elements of an ambient space. Mathematically any measure over a finite space can be characterized by N real numbers (Figure 8)

$$\mu = \{m_1, \dots, m_N\}$$

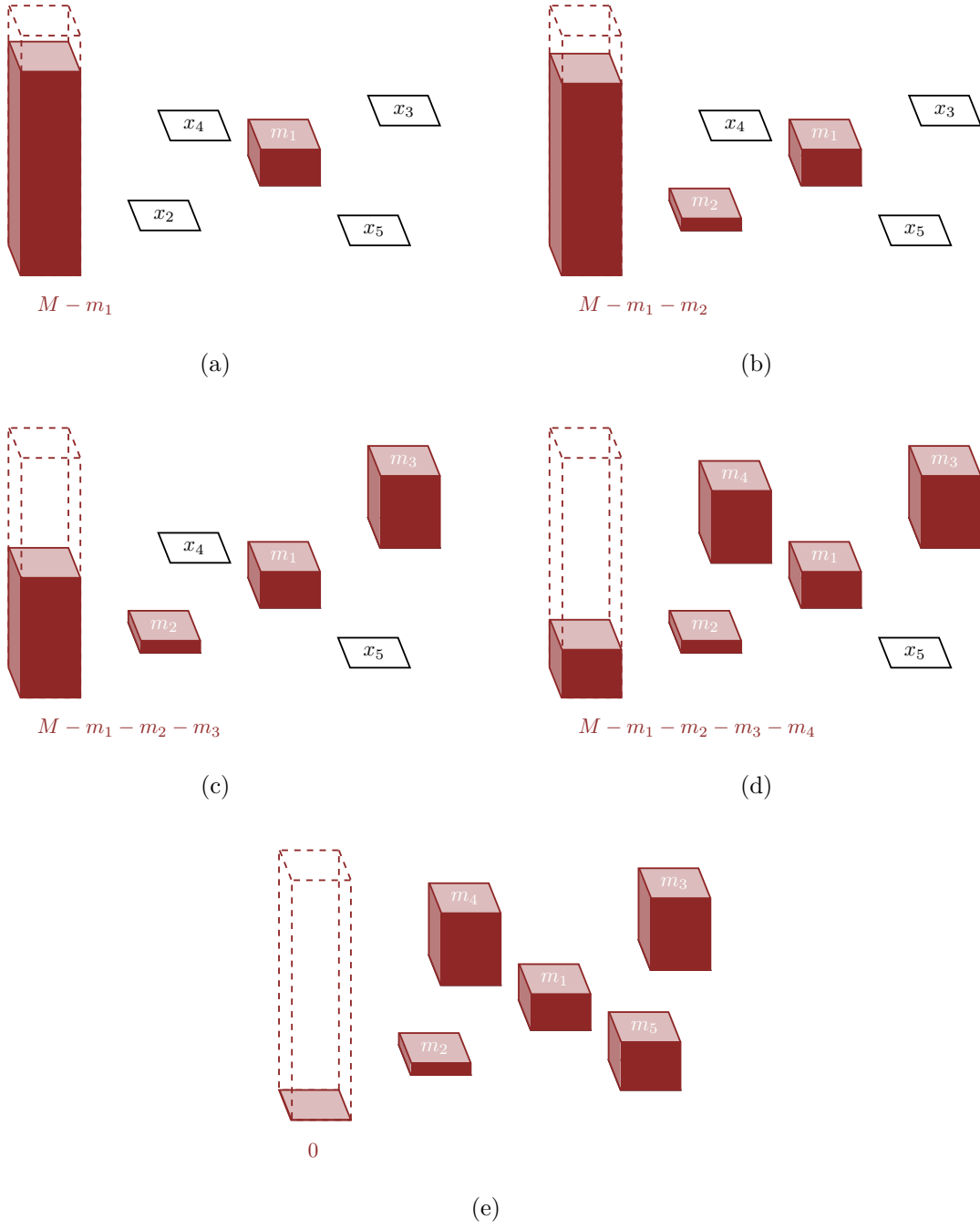


Figure 7: Because the total quantity M is conserved every allocation m_n to an element $x_n \in X$ depletes the amount available for the allocation to the remaining elements. An exhaustive allocation leaves nothing left in the initial reservoir after each element has received its allocation.

that satisfy

$$0 \leq m_n$$

and

$$\sum_{n=1}^N m_n = M.$$

Recall that M can be infinite.

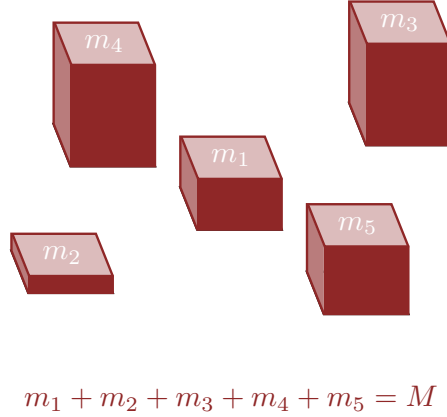


Figure 8: A measure μ over the finite space X is any consistent allocation of M to the elements $x_n \in X$. Every measure can be characterized by N real numbers m_n that sum to M , or equivalently a function that maps each x_n to m_n .

The larger m_n the more of M is allocated to the element x_n . Following this terminology we will also refer to M as the **total measure** and m_n as the **measure allocated to x_n** .

The allocation $\{x_n, m_n\}$ defined by a measure μ can also be organized into a function that maps each element to its associated allocation,

$$\begin{aligned} \mu : X &\rightarrow \mathbb{R}^+ \\ x_n &\mapsto m_n = \mu(x_n). \end{aligned}$$

A nice conceptual benefit of this function perspective is that instead of thinking about the global allocation m_1, \dots, m_N all at once we can reason about only local allocations by evaluating μ at a single input $m_n = \mu(x_n)$ one at a time.

Now there are an infinite number of ways to allocate a total measure to the elements of a finite space, and hence an infinite number of measures. I will denote the space of all possible measures over X as $\mathcal{M}(X)$.

Within this space there are a few notable examples. For example a **singular measure** allocates the total measure M to a single element, leaving the rest with nothing (Figure 9a). On the other hand a **uniform measure** allocates the same measure M/N to each element (Figure 9b). On finite spaces there are N distinct singular measures, one for each distinct element, and a single unique uniform measure.

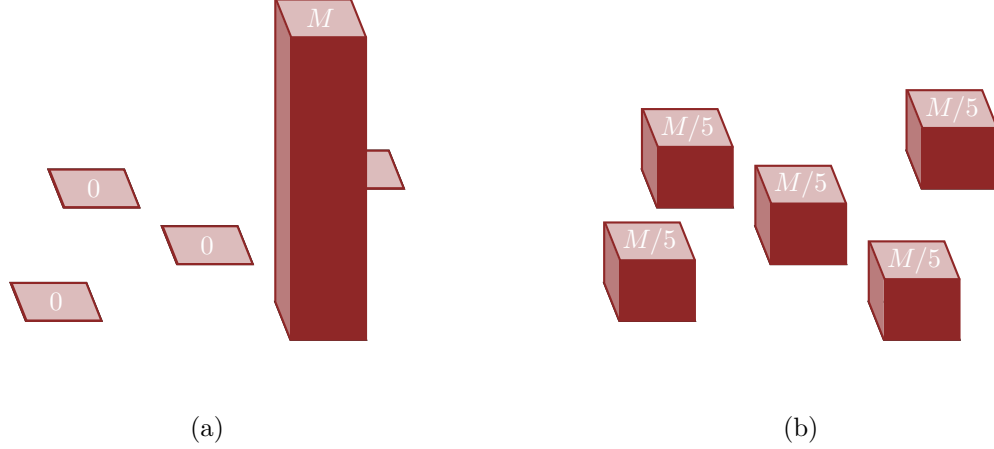


Figure 9: A singular measure (a) allocates the total measure to a single element while the uniform measure (b) spreads the total measure to each element evenly.

Perhaps the most important class of measures, however, are measures where the total measure is finite, $M < \infty$. Appropriately enough we refer to these measures as **finite measures**.

What makes finite measures so special is that we can always reframe the allocation they define into a *relative* one. Instead of considering the absolute measure allocated to each element m_n we can consider the *proportion* of the total measure allocated to each element, (Figure 10)

$$p_n = m_n / M.$$

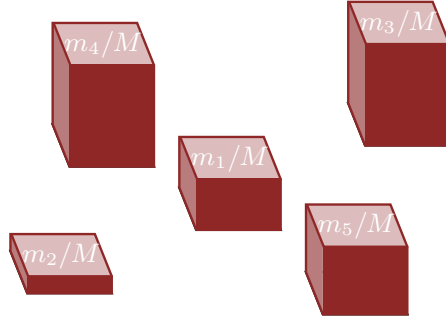
By construction proportions are confined to the unit interval $[0, 1]$. As with any quantity taking values in $[0, 1]$ we can represent proportions equally well with decimals, for example $p_n = 0.2$, and percentages, $p_n = 20\%$.

In other words a proportional measure defines the function (Figure 11)

$$\begin{aligned} \pi : X &\rightarrow [0, 1] \\ x_n &\mapsto p_n = \pi(x_n) \end{aligned}$$

with

$$0 \leq p_n \leq 1$$



$$\frac{m_1}{M} + \frac{m_2}{M} + \frac{m_3}{M} + \frac{m_4}{M} + \frac{m_5}{M} = \frac{M}{M} = 1$$

Figure 10: Every finite measure can be characterized by a proportional allocation.

and

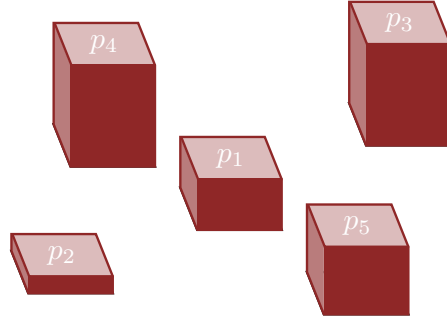
$$\sum_{n=1}^N p_n = 1.$$

A collection of variables $\{p_1, \dots, p_N\}$ satisfying these properties is also referred to as a **simplex**.

More importantly a proportional measure π is also known as a **probability distribution** with the proportional allocations p_n denoted **probabilities**. In other words while the term “probability” is often encumbered with all kinds of interpretational and philosophical baggage its mathematical structure is really quite straightforward. On a finite space a probability is just the proportion of some finite quantity that is allocated to an individual element.

The philosophical tension arises only when we try to assign some meaning to that quantity. This isn’t a question of mathematics but rather *applying* mathematics to model a system of interest. We’ll come back to the many different systems that can be consistently modeled with probability theory, and hence the interpretations of probability itself, in Chapter 4.

While we’re on the topic of terminological issues the word “distribution” is not without its own problems. In mathematics the term “distribution” is heavily overloaded and can be used to refer to a variety of related but distinct concepts. One way to avoid confusion is to always refer to proportional measures as “probability distributions” and never just “distributions” alone.



$$p_1 + p_2 + p_3 + p_4 + p_5 = 1$$

Figure 11: A proportional allocation is also known as a probability distribution.

3 Measure and Probability Over Subsets

On finite spaces any allocation, absolute or proportional, over the individual elements $x \in X$ also defines an allocation over entire subsets $\mathbf{x} \in 2^X$. The measure allocated to a subset is just the sum of the measures allocated to the elements in that subset. For example the measure allocated to $\mathbf{x} = \{x_1, x_2, x_4\}$ is $m_1 + m_2 + m_4$ (Figure 12).

In other words a measure over individual elements $\mu : X \rightarrow \mathbb{R}^+$ is equivalent to a measure over subsets $\mu : 2^X \rightarrow \mathbb{R}^+$. Similarly a probability distribution over individual elements $\pi : X \rightarrow [0, 1]$ is equivalent to a probability distribution over subsets $\pi : 2^X \rightarrow [0, 1]$. Note that we have made the cardinal sin of *overloading* our notation so that μ refers to both types of measures, and the only way to differentiate them is through their inputs; $\mu(x)$ is the point-wise measure allocated to $x \in X$ and $\mu(\mathbf{x})$ is the subset-wise measure allocated to $\mathbf{x} \in 2^X$.

By construction any subset measure and probability distribution satisfy a wealth of useful properties. For example for any measure

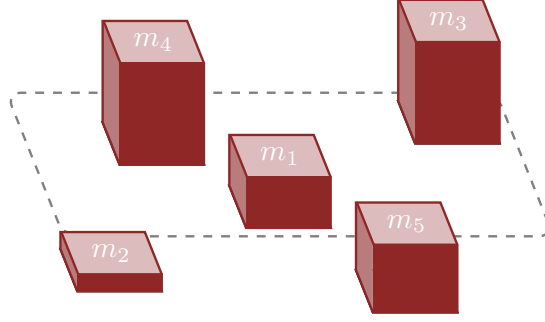
$$\mu(\emptyset) = 0$$

and

$$\mu(X) = \sum_{n=1}^N \mu(x_n) = M,$$

while for any probability distribution we have $\pi(\emptyset) = 0$ and $\pi(X) = 1$.

Even better the subset allocations play well with the subset operations. Consider for example the two disjoint subsets $\mathbf{x}_1 = \{x_1, x_3\}$ and $\mathbf{x}_2 = \{x_2, x_5\}$. Because the two subsets are disjoint



$$\mathbf{x} = \{x_1, x_3, x_4\}$$

$$\mu(\mathbf{x}) = m_1 + m_3 + m_4$$

Figure 12: On a finite space an allocation over individual elements also defines an allocation over subsets.

their union combines all of their elements,

$$\mathbf{x}_1 \cup \mathbf{x}_2 = \{x_1, x_2, x_3, x_5\}$$

and

$$\begin{aligned} \mu(\mathbf{x}_1 \cup \mathbf{x}_2) &= \mu(\{x_1, x_2, x_3, x_5\}) \\ &= m_1 + m_2 + m_3 + m_5 \\ &= (m_1 + m_3) + (m_2 + m_5) \\ &= \mu(\mathbf{x}_1) + \mu(\mathbf{x}_2). \end{aligned}$$

More generally for any collection of subsets

$$\mathbf{x}_1, \dots, \mathbf{x}_K$$

that are mutually disjoint,

$$\mathbf{x}_k \cap \mathbf{x}_{k'} = \emptyset$$

for $k \neq k'$, we have

$$\mu(\cup_{k=1}^K \mathbf{x}_k) = \sum_{k=1}^K \mu(\mathbf{x}_k).$$

In other words if we can decompose a subset into a disjoint collection of smaller subsets then we can decompose the measure allocated to that initial subset into measures allocated to the component subsets. This consistency property is known as **additivity**.

A subset x and its complement x^c always disjoint, $x \cap x^c = \emptyset$. At the same time their union covers the entire space, $x \cup x^c = X$. Consequently additivity implies that

$$\begin{aligned} M &= \mu(X) \\ &= \mu(x \cup x^c) \\ &= \mu(x) + \mu(x^c), \end{aligned}$$

or

$$\mu(x^c) = M - \mu(x).$$

In words that the measure allocated to the complement of a subset is the total measure less the measure allocated to that subset. For probability distributions this becomes even cleaner,

$$\pi(x^c) = 1 - \pi(x).$$

When two subsets overlap we have to take into consideration that the sum of their measures $\mu(x_1) + \mu(x_2)$ double counts the allocations to any elements shared between them. For example if $x_1 = \{x_1, x_4\}$ and $x_2 = \{x_1, x_5\}$ then

$$\begin{aligned} \mu(x_1 \cup x_2) &= \mu(\{x_1, x_4, x_5\}) \\ &= m_1 + m_4 + m_5. \end{aligned}$$

but

$$\begin{aligned} \mu(x_1) + \mu(x_2) &= (m_1 + m_4) + (m_1 + m_5) \\ &= m_1 + m_1 + m_4 + m_5 \\ &= m_1 + \mu(x_1 \cup x_2). \end{aligned}$$

The elements that is double counted here, however, is just the lone element in the intersection of the two subsets. Consequently the excess measure allocated to the union is given by (Figure 13)

$$m_1 = \mu(\{x_1\}) = \mu(x_1 \cap x_2).$$

and we can write

$$\mu(x_1) + \mu(x_2) = \mu(x_1 \cap x_2) + \mu(x_1 \cup x_2).$$

This relationship is quite general and holds for *any* two subsets regardless of their overlap.

Because of these subset properties we can build up a given measure in many different ways, each of which can be useful in different circumstances. This provides extremely convenient flexibility when trying to apply measure theory and probability theory in practice.

For example we can always specify a measure *globally* by specifying the individual allocations at the same time (Figure 14). Alternatively we can specify the allocation *locally* by considering each element one at a time (Figure 15). At each iteration we can take only from the measure allocated to the remaining elements, which corresponds to the “reservoir”.

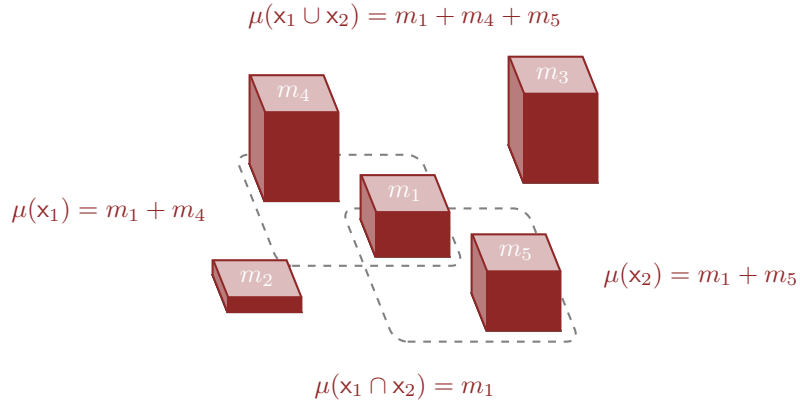


Figure 13: When two subsets overlap the measure allocated to each doubles the measure allocated to any overlapping elements, here x_1 , but the measure allocated to their union does not. This results in an important relationship between the measures allocated to the two subsets, the measure allocated to their union, and the measure allocated to their intersection.

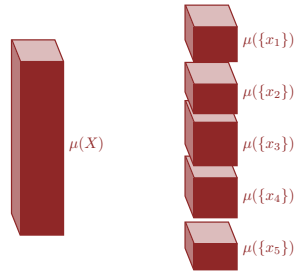


Figure 14: Measures can be constructed by specifying the individual element allocations all at once.

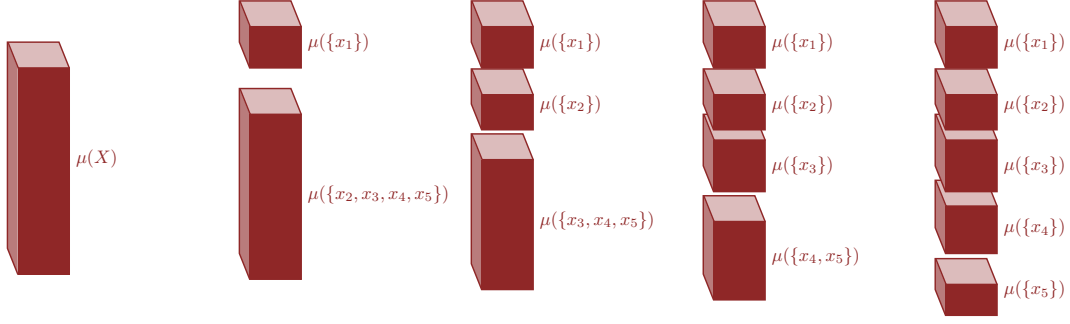


Figure 15: At the same time measures can be constructed by specifying the individual element allocations one by one.

Critically we do not always need to start with individual allocations. Instead we can always start by allocating the total measure to disjoint subsets and then iteratively *refining* that allocation to smaller and smaller subsets until we reach the individual elements (Figure 16).

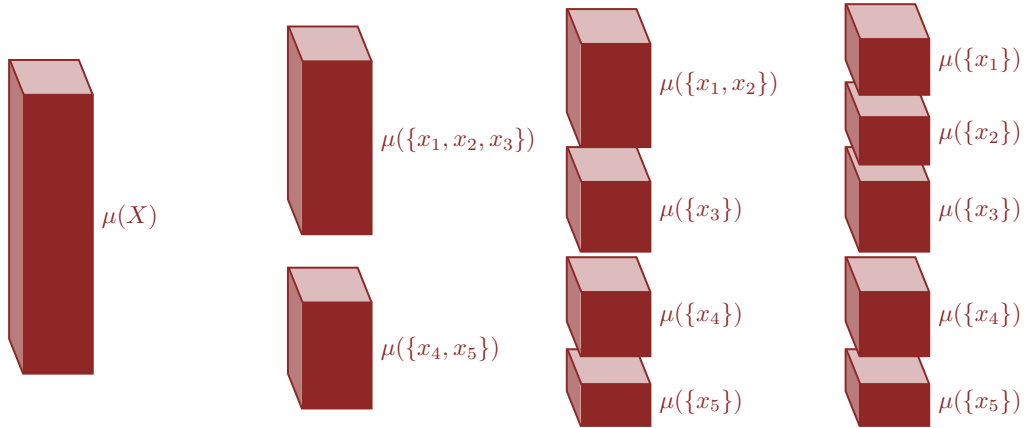


Figure 16: Finally measures can be constructed by allocating the total measure to disjoint subsets and then iteratively refining that allocation to smaller and smaller subsets.

In addition to providing this construction flexibility the subset definition of a measure $\mu : 2^X \rightarrow \mathbb{R}^+$ is also critical for generalizing measure theory beyond finite spaces. Specifically it becomes *necessary* when trying to consistently define measures on more mathematically complicated spaces like the real line. This will be the topic of Chapter 3.

4 Acknowledgements

A very special thanks to everyone supporting me on Patreon: Aapeli Nevala, Adam Bartonicek, Adam Fleischhacker, Adan, Adriano Yoshino, Alan Chang, Alessandro Varacca, Alexander Bartik, Alexander Noll, Alexander Rosteck, Anders Valind, Andrea Serafino, Andrew Mascioli, Andrew Rouillard, Andrew Vigotsky, Angie_Hyunji Moon, Ara Winter, asif zubair, Austin Rochford, Austin Rochford, Avraham Adler, Ben Matthews, Ben Swallow, Brynjolfur Gauti Jónsson, Cameron Smith, Canaan Breiss, Cat Shark, Charles Naylor, Chase Dwelle, Chris Zawora, Christopher Mehrvarzi, Chuck Carlson, Colin Carroll, Colin McAuliffe, Cruz, Damien Mannion, Damon Bayer, dan mackinlay, Dan Muck, Dan W Joyce, Dan Waxman, Dan Weitzenfeld, Daniel, Daniel Edward Marthaler, Daniel Rowe, Darshan Pandit, Darthmaluus, David Burdelski, David Galley, David Humeau, David Wurtz, dilsher singh dhillon, Doug Rivers, Dr. Jobo, Dr. Omri Har Shemesh, Ed Cashin, Ed Henry, Edgar Merkle, edith darin, Eric LaMotte, Erik Banek, Ero Carrera, Eugene O’Friel, Evan Cater, Fabio Pereira, Fabio Zotte, Felipe González, Fergus Chadwick, Finn Lindgren, Florian Wellmann, Francesco Corona, Geoff Rollins, George Ho, Granville Matheson, Greg Sutcliffe, Guido Biele, Hamed Bastan-Hagh, Haonan Zhu, Hector Munoz, Henri Wallen, Hugo Botha, Huib Meulenbelt, Håkan Johansson, Ian Costley, Ian Koller, idontgetoutmuch, Ignacio Vera, Ilaria Prosdocimi, Isaac S, Isaac Vock, J, J Michael Burgess, Jair Andrade, James Hodgson, James McInerney, James Wade, Janek Berger, Jason Martin, Jason Pekos, Jeff Burnett, Jeff Dotson, Jeff Helzner, Jeffrey Burnett, Jeffrey Erlich, Jesse Wolfhagen, Jessica Graves, Joe Wagner, John Flournoy, Jon, Jonathan H. Morgan, Jonathan St-onge, Jonathon Vallejo, Joran Jongerling, Joseph Despres, Josh Weinstock, Joshua Duncan, Joshua Griffith, Josué Mendoza, JU, Justin Bois, Karim Naguib, Karim Osman, Keith O’Rourke, Kejia Shi, Kevin Foley, Kristian Gårdhus Wichmann, Kádár András, lizzie, LOU ODETTE, Luiz Pessoa, Marc Dotson, Marc Trunjer Kusk Nielsen, Marcel Lüthi, Marek Kwiatkowski, Mark Donoghoe, Mark Worrall, Markus P., Martin Modrák, Matt Moores, Matthew, Matthew Kay, Matthieu LEROY, Maurits van der Meer, Merlin Noel Heidemanns, Michael Colaresi, Michael DeWitt, Michael Dillon, Michael Lerner, Mick Cooney, Márton Vaitkus, N Sanders, Name, Nathaniel Burbank, Nic Fishman, Nicholas Clark, Nicholas Cowie, Nick S, Nicolas Frisby, Octavio Medina, Ole Rogeberg, Oliver Crook, Olivier Ma, Pablo León Villagrà, Palwasha Khan, Patrick Kelley, Patrick Boehnke, Pau Pereira Batlle, Peter Smits, Pieter van den Berg, ptr, Putra Manggala, Ramiro Barrantes Reynolds, Ravin Kumar, Raúl Peralta Lozada, Riccardo Fusaroli, Richard Nerland, Robert Frost, Robert Goldman, Robert kohn, Robert Mitchell V, Robin Taylor, Rong Lu, Ross McCullough, Ryan Grossman, Rémi, S Hong, Sam Levy, Scott Block, Scott Brown, Sean Pinkney, Sean Wilson, Serena, Seth Axen, shira, Simon Duane, Simon Lilburn, Srivatsa Srinath, sssz, Stan_user, Stefan, Stephanie Fitzgerald, Stephen Lienhard, Steve Bertolani, Stone Chen, Sus, Susan Holmes, Svilup, Sören Berg, Tagir Akhmetshin, Tao Ye, Tate Tunstall, Tatsuo Okubo, Teresa Ortiz, Thiago de Paula Oliveira, Thomas Lees, Thomas Vladeck, Tiago Cabaço, Tim Radtke, Tom McEwen, Tomáš Frýda, Tony Wuersch, Virginia Fisher, Vitaly Druker, Vladimir Markov, Wil Yegelwel, Will Farr, Will Kurt, Will Tudor-Evans, woejozney, yolhaj, yureq, Zach A, Zad Rafi, and Zhengchen Cai.

License

A repository containing all of the files used to generate this chapter is available on [GitHub](#).

The text and figures in this chapter are copyrighted by Michael Betancourt and licensed under the CC BY-NC 4.0 license:

<https://creativecommons.org/licenses/by-nc/4.0/>