

Introduction to Measure Theoretic Probability

Sam Leone

December 27, 2022

Chapter 1

Motivation & Preliminaries

Why study probability theory? If you're anything like me, you know the basics: dice rolls, taking expectations, and some basic distributions. But along the way, you had a few lingering questions: Why is the law of large numbers true? The central limit theorem? What the heck is a probability density, really? What does it even mean to condition on something with probability 0? There are also the existence of conditioning paradoxes. Perhaps you've heard about stochastic processes and the different limit theorems which show that random processes converge to an equilibrium. How the heck can you show that? We address these through a fundamental shift in perspective: remove the randomness from the study of probability. Rather than studying the "probability" of an event, study its size. The simple idea to use tools from measure theory has far reaching consequences. The added trouble may make you think — why are we doing this? But after a few headaches, the applications will be well worth it.

We will assume a very basic knowledge of real analysis.

1.1 Events, Sizes & The Universe

Let's begin our study of probability with a classic example: the roll of a fair die. We all know that *each side of a die occurs with probability $1/6$* . But what do we even mean by this? Some people think of it from a frequentist perspective — if you roll the die 600 billion billion times, it will come up 2 about 100 billion billion times (assuming it doesn't break). In a way, this is silly and circular: we're defining probabilities by the Law of Large Numbers? What could we possibly extend this to densities? If we throw a dart 100 billion billion times, it will most likely hit a given spot 0 times.

The more prudent & careful approach is to think about the universe of possible events, and assign sizes to suitable subsets of those events. Formally, the universe is given by Ω . In the case of the dice roll, you could think of Ω as

describing every outcome in every instantiation of this dice roll in the multiverse. The amount of information captured by Ω could be arbitrary - it could contain the outcome of the dice roll, the weather, and what you get for Christmas. But for our game of Monopoly, we don't really care about *everything*. Hence we consider a family of subsets of interest \mathcal{F} . For example, \mathcal{F} might consist of the event a 1 is rolled, a 2 is rolled, and all combinations of these. Even if the set of possibilities where a 1 is rolled can be further split up by weather, we essentially turn a blind eye to these distinctions. In our case, \mathcal{F} is called a σ -algebra ("sigma algebra") and has the sensible closure properties.

Definition 1 (σ -algebra). *A family of subsets \mathcal{F} of Ω is called a σ algebra if,*

- $\emptyset \in \mathcal{F}, \Omega \in \mathcal{F}$
- *Closure under complement: For all $A \in \mathcal{F}$, $A^c \in \mathcal{F}$ as well*
- *Closure under countable union: If $\{A_i\}_{i \in I}$ is a countable set such that $A_i \in \mathcal{F}$ for all $i \in I$, then $\cup_{i \in I} A_i \in \mathcal{F}$ as well*

One can verify that these properties imply σ -algebras are also closed under countable union. These requirements are quite natural when considering the operations we normally do in probability.

Finally, we need a machine which computes sizes. This is done through a so-called *measure* μ . So in the dice roll, $\{\text{roll a 6}\} \in \mathcal{F}$, and $\mu(\{\text{roll a 6}\}) = 1/6$. Formally, μ can be regarded as a set $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$. Note that, in the probabilistic case, $\mu(\Omega) = 1$ (the size of everything is 1), but this need not be true in general. In fact, we will consider a notable exception: the Lebesgue measure. The necessary properties of μ pair nicely with the definition of a σ -algebra. In short, we require nonnegative sizes, the size of nothing to be 0, and that the sizes of non-overlapping things adds.

Definition 2 (Measure). *A measure on \mathcal{F} is a function $\mu : \mathcal{F} \rightarrow \mathbb{R}$ such that,*

- *For all $A \in \mathcal{F}$, $\mu(A) \geq 0$*
- $\mu(\emptyset) = 0$
- *If $\{A_i\}_i$ are countable in \mathcal{F} and pairwise disjoint, then $\mu(\cup_{i \in I} A_i) = \sum_{i \in I} \mu(A_i)$*

In particular, if $\mu(\Omega) = 1$, μ is a *probability measure*. Hopefully, the first two conditions are clear and well-motivated. For the last one, we are simply requiring something like $\mu(\{\text{roll a 1} \cup \{\text{roll a 2}\}\}) = \mu(\{\text{roll a 1}\}) + \mu(\{\text{roll a 2}\})$. There are a few special cases worth familiarizing ourselves with: probability measures \subseteq finite measures \subseteq σ -finite measures:

Definition 3 (Finite Measure). *If $\mu(\Omega) < \infty$, then μ is finite.*

Definition 4 (Probability Measure). *If $\mu(\Omega) < \infty$, then μ is a probability measure.*

Definition 5 (σ -finite). If $\Omega = \cup_{i \in I} A_i$, such that I is countable and each $A_i \in \mathcal{F}$ but $\mu(A_i) < \infty$, then μ is σ -finite

Definition 6 (Measure Space). A measure space is a triple $(\Omega, \mathcal{F}, \mu)$ where \mathcal{F} is a σ -algebra of subsets of Ω , and μ is a measure on \mathcal{F}

1.2 σ -algebras and Generating Sets

Suppose \mathcal{E} is a family of subsets of Ω (now, we make no assumptions on the nature of \mathcal{E}). We say the σ -algebra generated by \mathcal{E} , denoted $\sigma(\mathcal{E})$ is the smallest σ -algebra containing \mathcal{E} :

$$\sigma(\mathcal{E}) = \bigcap_{\sigma\text{-algebras } \mathcal{F} \text{ s.t. } \mathcal{E} \subseteq \mathcal{F}} \mathcal{F}$$

In this sense, \mathcal{E} can be thought of as the atoms of Ω from which we build molecules in \mathcal{F} . As we will see, these atoms need not be unique. For example, if $\Omega = \{1, 2, 3, 4, 5, 6\}$, if $\mathcal{E} = \{\{1\}, \{2\}, \{3\}, \dots, \{6\}\}$, then $\sigma(\mathcal{E}) = \mathcal{P}(\Omega)$, the full power set. However, if $\mathcal{E} = \{\{1, 2, 3\}, \{4, 5, 6\}\}$, then the resulting structure has a “lower resolution:” $\sigma(\mathcal{E}) = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}$. Note that the following intuitive properties hold:

Lemma 1.2.1. Let $\mathcal{E} \subseteq \mathcal{P}(\Omega)$. Then,

- If \mathcal{E}_1 is a σ -algebra, then $\sigma(\mathcal{E}) = \mathcal{E}$
- If $\mathcal{E} \subseteq \mathcal{E}'$, then $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$

Proof. Clearly, as \mathcal{E} is a σ -algebra containing \mathcal{E} ,

$$\sigma(\mathcal{E}) = \mathcal{E} \cap \bigcap_{\sigma\text{-algebras } \mathcal{F} \text{ s.t. } \mathcal{E} \subseteq \mathcal{F}} \mathcal{F} \subseteq \mathcal{E}$$

Also, $\sigma(\mathcal{E}) \supseteq \mathcal{E}$ by definition. We conclude $\sigma(\mathcal{E}) = \mathcal{E}$. For the latter claim, we prove $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$ by considering arbitrary elements of $\sigma(\mathcal{E})$. Suppose $A \in \sigma(\mathcal{E})$. By definition, for all \mathcal{F} containing \mathcal{E} , $A \in \mathcal{F}$. Note also that for all \mathcal{F}' containing \mathcal{E}' , \mathcal{F}' contains \mathcal{E} as well, so $A \in \mathcal{F}'$. As a consequence $A \in \sigma(\mathcal{E}')$. Since A was generic, $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$. □

1.2.1 The Borel σ -algebra

It would be no overstatement to say the most often studied σ -algebra is the Borel σ -algebra. In this case, $\Omega = \mathbb{R}$, \mathcal{E} = the open sets in \mathbb{R} , and $\mathcal{F} = \sigma(\mathcal{E})$. This is denoted $\mathcal{B}(\mathbb{R})$. More generally, the Borel σ -algebra of a metric space \mathcal{X} is denoted $\mathcal{B}(\mathcal{X})$. Recall from definition 1 that $\mathcal{B}(\mathbb{R})$ should be closed under compliment, and so $\mathcal{B}(\mathbb{R})$ contains the closed sets as well. Moving forward, the Borel σ -algebra will contain all the richness we will practically need.

1.2.2 Lebesgue-Stieltjes Measures

A generic class of measures is the set of Riemann-Stieltjes Measures. A distribution function (think CDF) is a map $F : \mathbb{R} \rightarrow \mathbb{R}$ such that,

- $F(x)$ is nondecreasing with x
- F is right continuous ($\lim_{y \rightarrow x^+} F(y) = F(x)$)

The corresponding Lebesgue-Stieltjes measure sets $\mu((a, b]) = F(b) - F(a)$. It can be shown that this is enough to specify the whole measure over $\mathcal{B}(\mathbb{R})$.

Example 1 (The Lebesgue Measure). *The Lebesgue measure is the Lebesgue-Stieltjes Measure when $F(x) = x$, and so $\mu((a, b]) = b - a$.*

Example 2 (The Normal Distribution). *The normal distribution is induced by the Lebesgue-Stieltjes measure with $\mu((a, b]) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.*

Example 3 (Probability Mass Functions from CDFs). *More generally, when F is the CDF of a probability mass function, the corresponding Lebesgue-Stieltjes measure corresponds to that probability distribution.*