Probability with Elementary Measure Theory

Leonardo T. Rolla

2021-08-05

Contents

Preface

Reusing this material

Source code

If you make changes to these notes, please distribute the modified source code along with the typeset material, and please keep this request to those who receive it. This is a kind request from the author, not a legal requirement.

Attribution

This is an updated and expanded version of the lecture notes “Mathematics of Random Events” written in collaboration with Nikolaos Constantinou in 2020. Those notes were largely based on material typeset by Nikolaos Constantinou as a student, when the ST342 module was taught by Wei Wu in the Autumn 2019 term at Warwick. The module taught by Wei Wu was largely based on handwritten lecture notes received from Larbi Alili, which in turn were based on material previously developed by Sigurd Assing, Anastasia Papavasileiou, Jon Warren and Wilfrid Kendall.

Chapter 1

Infinity

Can we toss a coin infinitely many times?

Why is countable additivity important?

We know that $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, but why? It doesn't seem to follow from the usual definition that treats discrete and continuous random variables as hermetically separated entities. So, what is expectation, really?

Do continuous random variables even exist?

How small can a class $\mathcal{A}$ of events be so that their probabilities determine $\mathbb{P}$?

Why can we differentiate moment generating functions to compute moments?

Our aim is to study concepts of Measure Theory useful to Probability Theory, providing a solid ground for the latter. A *measure* generalises the notion of area for arbitrary sets in Euclidean spaces $\mathbb{R}^d$, $d \geq 1$. We introduce the theory of measurable spaces, measurable functions, integral with respect to a measure, density of measures, product measures, and convergence of functions and random variables.

Below we give examples that motivate the need for such a theory, discuss in which sense modern Measure Theory is the best we can hope for, and introduce the concept of infinite numbers and infinite sums used throughout the remaining chapters.

1.1 Events at infinity

We know that $A_n \uparrow A$ implies $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ which, assuming that $\mathbb{P}$ is non-negative and finitely additive, is equivalent to $\mathbb{P}$ being σ-additive (a shorthand for countably additive). Below we see examples where interesting models and events require some sort of limiting process in their study.

Example 1.1 (Ruin Probability)**.** Gambling with initial wealth $X_0 \in \mathbb{N}_0$. For any $t \geq 1$, we bet an integer amount and reach a wealth denoted by X_t. If at any point in time, wealth amounts to 0, it remains 0 forever. The sample space that indicates the wealth process is $\Omega = \{(x_0, x_1, \dots) : x_i \in \mathbb{N}_0\}$. We define the function of wealth after the n-th gamble by, $X_n : \Omega \to \mathbb{N} \cup \{0\}$, where $X_n(\omega) = x_n$. In fact, $(X_n)_{n \geq 0}$ is a Markov process. Then

$$\{\text{stay in state 0 eventually}\} = \bigcup_{n=0}^{\infty} \bigcap_{m=n}^{\infty} \{\omega \in \Omega : X_m(\omega) = 0\}.$$

By monotonicity and continuity, its probability is

$$\mathbb{P}(\{\text{stay in state 0 eventually}\}) = \lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : X_n(\omega) = 0\}).$$

Example 1.2 (Brownian motion)**.** It is possible to construct a sequence of continuous piece-wise linear functions, which in the limit give a continuous nowhere differentiable random path that is at the core of Stochastic Analysis.

Example 1.3 (Uniform variable)**.** If $X \sim U[0,1]$, then $\mathbb{P}(X \neq x) = 1$ for every $x \in \mathbb{R}$. Nevertheless, $\mathbb{P}(X \neq x \text{ for every } x \in \mathbb{R}) = 0$, so there will be some unlucky x that will happen to be hit by X. Now for a countable set A, by σ-additivity we have $\mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x) = 0$. Since $\mathbb{Q}$ is countable, we see that a uniform random variable is always irrational. Well, this is unless there is an uncountable number of such variables in the same probability space, in which case some unlucky variables may happen to take rational values.

Example 1.4. Let $X_1, X_2, X_3, \dots$ be independent and take value ± 1 with probability $\frac{1}{2}$ each, and take $S_n = X_1 + \cdots + X_n$. The strong law of large numbers says that, almost surely (a shorthand for "$\mathbb{P}(\cdots) = 1$"), $\frac{S_n}{n} \to 0$ as $n \to \infty$. In order to show this, we need to express this event as a limit and compute its probability. The first part is simply:

$$\{\lim_{n\to\infty} \tfrac{S_n}{n} = 0\} = \bigcap_{k=1}^{\infty} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{|\tfrac{S_n}{n}| < \tfrac{1}{k}\}.$$

Example 1.5 (Recurrence of a random walk)**.** For the sequence $(S_n)_n$ of the previous example, we know that $\frac{S_n}{n} \to 0$ with probability one. We want to consider whether $\frac{S_n}{n}$ converges to zero from above, from below, or oscillating, which is the same as asking whether $S_n = 0$ infinitely often.

1.2 The measure problem

It is clear (it should be!) from previous examples that we want to work with measures that have nice *continuity* properties, so we can take limits. However, when the mass is spread over uncountably many sample points $\omega \in \Omega$, it is not possible to assign a measure to all subsets of Ω in a reasonable way.

We would like to define a random variable uniformly distributed on $[0, 1]$, by means of a function that assigns a weight to subsets of this interval. For instance, what is the probability that this number is irrational? What is the probability that its decimal expansion does not contain a 3?

This is the same problem as assigning a 'length' to subsets of $\mathbb{R}$. We are also interested in defining a measure of 'area' on $\mathbb{R}^2$, 'volume' on $\mathbb{R}^3$, and so on.

Of course, a good measure of length/area/volume/etc. on $\mathbb{R}^d$ should:

1. give the correct value on obvious sets, such as intervals and balls;
2. give the same value if we rotate or translate a set;
3. be σ-additive.

We stress again that σ-additivity is equivalent to a measure being continuous, and we are not willing to resign that. On the other hand, we do not want more than that: each of the uncountably many points in $[0, 1]$ alone has length zero, but all together they have length one; likewise, each sequence of coin tosses in $\{0, 1\}^{\mathbb{N}}$ has probability zero, but all together they have probability one.

The measure problem is the following.

> There is no measure m defined on all subsets of $\mathbb{R}^d$ which satisfy all the reasonable properties listed above. What modern Measure Theory does is to work with measures that are defined on a class of sets which is large enough to be useful and small enough for these properties to hold.

The next example shows that there is no measure m defined on all subsets of $\mathbb{R}^3$ which satisfies these three properties.

Example 1.6 (Banach-Tarski paradox)**.** Consider the ball

$$B = \{x \in \mathbb{R}^3 : \|x\| \leq 1\}.$$

There exist[1] $k \in \mathbb{N}$, *disjoint* sets $A_1, \dots, A_{2k}$, and *isometries* (maps that preserve distances and angles) $S_1, \dots, S_{2k} : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$B = (A_1 \cup \dots \cup A_k) \cup (A_{k+1} \cup \dots \cup A_{2k}),$$

$$B = S_1 A_1 \cup \dots \cup S_k A_k, \quad B = S_{k+1} A_{k+1} \cup \dots \cup S_{2k} A_{2k}.$$

[1]See https://youtu.be/s86-Z-CbaHA for a nice overview of the proof.

So B was decomposed into *finitely many* pieces, which were later on moved around *rigidly* and *recombined* to produce two copies of B! Why is it a paradox? Finitely many pieces is not the issue in itself, since $\mathbb{N}$ can be decomposed into even and odd numbers, and they can be *compressed* (or stretched, in some sense) to produce two copies of $\mathbb{N}$. Rigidity alone is not the issue either, since we can move each of the *uncountably many* points of the segment $[0, 1]$ to form the segment $[0, 2]$. The paradox is that this magic was done with *rigid movements on finitely many pieces*. And here we can see the *measure problem*: if all these disjoint sets $A_1, \dots, A_{2k}$ were to have a volume $V_1, \dots, V_{2k} \geq 0$, what would be the volume of the ball B?

The next example is not nearly as effective in impressing friends at a party, and would certainly not make a youtube video with 31 million views, but it has two advantages. First, it shows directly that the measure problem already occurs on $d = 1$. Second, we can actually explain its proof in a third of a page rather than a dozen.

Example 1.7 (Vitali Set)**.** Consider the unit circle $\mathbb{S}^1$ with points indexed by turns instead of degrees or radians. This is the same as the interval $\mathbb{S}^1 = [0, 1)$ with the angle addition operation $x \oplus y = x + y \mod 1$. There exists a set $E \subseteq [0, 1)$ such that $\mathbb{S}^1$ is decomposed into disjoint $\{E_n\}_{n\in\mathbb{N}}$ which are translations of E. And here we see again the *measure problem*: by σ-additivity, if the length of E is zero, then the length of the circle is zero; and if the length of E is non-zero, then the length of the circle is infinite. So E is not measurable.

Proof. (Sketch) Write $\mathbb{Q} \cap [0, 1) = \{r_n\}_{n=1,2,3,\dots}$. For $E \subseteq \mathbb{S}^1$, let $E_n = \{x \oplus r_n : x \in E\}$ be the *translation* of E by r_n. We want to find a set E such that

1. The sets $E_1, E_2, E_3, \dots$ are disjoint,
2. The union satisfies $\cup_n E_n = \mathbb{S}^1$.

Start with a small set that satisfies the first property, such as $E = \{0\}$. Enlarge the set E by adding a point $x \in \mathbb{S}^1 \ \ (\cup_n E_n)$. Adding such point does not break the first property (proof omitted), and may help the second one. Keep adding points this way, until it is no longer possible. When it is no longer possible, it can only be so because the second property is also satisfied.[2] □

Remark. It is often emphasised that the Banach-Tarski paradox and Vitali set depend crucially on the Axiom of Choice (for the above sketch of proof, it is concealed in the expression "keep adding until"). We may wonder what happens if we do not accept this axiom. In this case, we cannot prove the Banach-Tarski paradox, nor the existence of a Vitali set. But neither can we prove that they do not exist, so the measure problem persists.

[2]See [**?**, 1.4.9] for a complete proof.

1.3 Infinite numbers and infinite sums

We now define the set of extended real numbers and briefly discuss some its useful properties, then discuss the meaning of infinite sums, and move on to other perhaps philosophical questions about this theory.

1.3.1 Extended real numbers

We are about to start working with measures, and because measures can be infinite, and integrals can be negative infinite, we work with the set of *extended real numbers* $\overline{\mathbb{R}} := [-\infty, +\infty]$ that extends $\mathbb{R}$ by adding two symbols $\pm\infty$. The novelty is of course to conveniently allow operations and comparisons involving these symbols.

Basically, we can safely operate as one would reasonably guess:

$$-\infty < -1 < 0 < 5 < +\infty,\ -7 + \infty = +\infty,\ (-2) \times (-\infty) = +\infty,$$
$$|-\infty| = +\infty,\ (+\infty) \times (-\infty) = -\infty,\ a \leq b \implies a + x \leq b + x$$
$$\lim_{n\to\infty}(2 + n^2) = \lim_{n\to\infty} 2 + \lim_{n\to\infty} n^2 = 2 + \infty = +\infty,$$

etc. Since we will never need to divide by infinity, let us leave $\frac{x}{\infty}$ undefined (otherwise we would need to check that x is finite).

The non-obvious definition is $0 \cdot \infty = 0$. In Calculus, it would have been considered an indeterminate form, but in Measure Theory it is convenient to define it this way because the integral of a function that takes value 0 on an interval of infinite length and $+\infty$ at a few points should still be 0. That is, the area of a rectangle having zero width and infinite length is zero.

Now some caveats. First, $\lim_n(a_n b_n) = (\lim_n a_n)(\lim_n b_n)$ may fail in case it gives $0 \cdot \infty$. Also, note that now $a + b = a + c$ does not imply $b = c$. This can be false when $a = \pm\infty$. Likewise, $a < b$ no longer implies that $a + x < b + x$. So we should be careful with cancellations.

> The one thing that is definitely not allowed, and that Measure Theory does not handle well, is
>
> $$\text{"}\ +\infty - \infty\ \text{"}\ !$$
>
> This is simply forbidden, and if we will ever write this, it will be in quotation marks and just in order to say that this case is excluded.

The reader should consult [**?**, §§B.4–B.6] and [**?**, p. xi] for a more complete description of operations on $\overline{\mathbb{R}}$.

1.3.2 Infinite sums

Infinite sums of numbers on $[0, +\infty]$ are always well-defined through a rather simple formula. If Λ is an index set and $x_\alpha \in [0, +\infty]$ for all $\alpha \in \Lambda$, we define:

$$\sum_{\alpha \in \Lambda} x_\alpha = \sup_{\substack{A \subseteq \Lambda \\ A \text{ finite}}} \sum_{\alpha \in A} x_\alpha.$$

The set Λ can be uncountable, but the sum can be finite only if $\Lambda_+ = \{\alpha : x_\alpha > 0\}$ is countable (proof omitted). If $x_\alpha \in [-\infty, +\infty]$, we define $\Lambda_- = \{\alpha : x_\alpha < 0\}$ and

$$\sum_{\alpha \in \Lambda} x_\alpha = \sup_{\substack{A \subseteq \Lambda_+ \\ A \text{ finite}}} \sum_{\alpha \in A} x_\alpha \; - \; \sup_{\substack{A \subseteq \Lambda_- \\ A \text{ finite}}} \sum_{\alpha \in A} -x_\alpha, \tag{1.1}$$

as long as this difference does not give "$+\infty - \infty$" !

The theory of conditionally convergent sums as

$$\sum_{j \in \mathbb{N}} x_j = \lim_n \sum_{j=1}^n x_j \tag{1.2}$$

is hardly meaningful to us. In case the expression in (1.1) is well-defined, we can write $(\Lambda_- \cup \Lambda_+) = \{\alpha_j\}_{j \in \mathbb{N}}$ by ordering these indices in any way we want (assuming for simplicity that these sets are countable), and formula (1.2) will give the same result as (1.1). Pretty robust.

However, in case (1.2) converges but (1.1) is not well-defined (so it gives "$+\infty - \infty$"), we can re-order the index set $\mathbb{N}$ so that (1.2) will give any number we want. This is definitely not the type of delicacy we want to handle here.

For this reason, we will only use (1.2) when either

$$x_j \in [0, +\infty]$$

for all j, or when

$$\sum_j |x_j| < \infty.$$

So there are two overlapping cases where we can work comfortably: non-negative extended numbers, or series which are absolutely summable.

1.3.3 Two different and overlapping theories

The above tradeoff is already a good prelude to something rather deep that will appear constantly in upcoming chapters. Borrowing from [**?**]:

> Because of this tradeoff, we will see two overlapping types of measure and integration theory: the *non-negative* theory, which involves quantities taking values in $[0, +\infty]$, and the *absolutely integrable* theory, which involves quantities taking values in $\mathbb{R}$ or $\mathbb{C}$.

However, at the risk of leaving it for the reader to figure out some corner cases, we can (and will) extend these theories to a theory on $[-\infty, +\infty]$ by doing what we just did above. Namely, whereas the absolutely integrable theory requires that both terms in (1.1) be finite, and the non-negative theory requires that one of them be zero, we only require that one of them be finite.

We end this chapter with the simplest example of these overlapping theories

Theorem 1.1 (Tonelli Theorem for series)**.** *Let $x_{m,n} \in [0, +\infty]$ be a doubly-indexed sequence. Then*

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} x_{m,n} = \sum_{(m,n)\in\mathbb{N}^2} x_{m,n} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} x_{m,n}.$$

Proof. A proof is given in §7.2 as an application of Tonelli Theorem, but here is a bare hands proof. For $A \subseteq \mathbb{N}^2$ finite, there exist $m, n \in \mathbb{N}$ such that $\sum_A x_{j,k} \leq \sum_{j=1}^n \sum_{k=1}^n x_{j,k}$. Conversely, given $m, n \in \mathbb{N}$, there is $A \subseteq \mathbb{N}^2$ such that $\sum_A x_{j,k} = \sum_{j=1}^n \sum_{k=1}^n x_{j,k}$. Therefore,

$$\sum_{(j,k)\in\mathbb{N}^2} x_{j,k} = \sup_A \sum_{(j,k)\in A} x_{j,k} = \sup_{m,n\in\mathbb{N}} \sum_{j=1}^{m}\sum_{k=1}^{n} x_{j,k} =$$
$$= \sup_{m\in\mathbb{N}} \sup_{n\in\mathbb{N}} \sum_{j=1}^{m}\sum_{k=1}^{n} x_{j,k} = \sup_{m\in\mathbb{N}} \sum_{j=1}^{m} \Big(\sup_{n\in\mathbb{N}} \sum_{k=1}^{n} x_{j,k} \Big) = \sum_{j=1}^{\infty} \Big(\sum_{k=1}^{\infty} x_{j,k} \Big).$$

The other equality is proved in identical way. □

Theorem 1.2 (Fubini Series)**.** *Let $x_{m,n} \in [-\infty, +\infty]$ be a doubly-indexed sequence. If*

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |x_{m,n}| < \infty,$$

then

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} x_{m,n} = \sum_{(m,n)\in\mathbb{N}^2} x_{m,n} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} x_{m,n}.$$

Proof. Given in §7.2 as an application of Fubini Theorem. □

Example 1.8. Consider the doubly-indexed sequence

$$x_{m,n} = \begin{array}{|cccccc} \hline 1 & -1 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 0 & 1 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{array}$$

which is not absolutely summable. Note that summing columns and then rows we get 1, whereas summing rows and then columns we get 0.

1.4 Dummy subsection

This section is to show how we cen refer to other sections and theorems.

Referencing the Theorems, exmalpes and equations is done by using their labels It is enough to enter `\@ref(label)` e.g.:

- Theorem 1.2 is a very important one.
- Don't forget to solve Example 1.7 before the exam.
- Equation (1.2) can be quite tricky.

To reference a section by its number we can do it the same way with `\@ref(label)`, e.g.:

- In Section 1.1 we introduce the topic of …

To reference a section by its title we can do this by witting the section header in square brackets `[Section name]`, e.g.:

- In Section Events at infinity we introduce the topic of …

Cross-references still work even when we refer to an item that is not on the current page of the HTML output.