Probability notes

Matthew Signorotti

Fall 2019

Contents

1 Sets and events

Most of this section, except for a brief excursion into formal axioms, is concerned with naive (intuitive) set theory, which appeals best to intuition rather than formal rigor. To start off, a **set** is a collection of abstract mathematical objects. **Subsets** A, B, … of some greater set Ω are also sets, but all of whose elements are contained in Ω. Note that the symbol Ω is often assumed, without specification otherwise, to be this “overarching” set over which subsets are taken. Moreover, the **empty set** contains no elements and is denoted $\emptyset$.

1.1 Functions and indicator functions

Functions are rules “mapping” from a set X to another (possibly identical) set Y. Give me some element of an input set, and a function relates some output element, deterministically, to this input. **Indicator functions** are functions from Ω to the binary integers $\mathbb{B} = \{0, 1\}$.

1.2 Set operations and axioms

1.2.1 Set operations

- The **complement** A^c of a subset A of Ω, with respect to Ω, is all elements $\omega \in \Omega$ such that $\omega \notin A$. (In math notation, this is $\{\omega : \omega \in \Omega, \omega \notin A\}$. The $\{\}$ indicates a set definition.)

- The **intersection** $AB = A \cap B$[1] of two sets A and B is all elements in both A and B ($\{\omega : \omega \in A, \omega \in B\}$). This definition extends naturally to longer or even arbitrarily long (infinite) sequences of sets.

- The **union** $A \cup B$ of two sets A and B is all elements in *either A or B* ($\{\omega : \omega \in A \text{ or } \omega \in B\}$). Again, this definition extends as expected to longer or infinite sequences of sets.

- The **set difference** $A \setminus B$ with respect to an overall set Ω is defined as

$$A \setminus B = AB^c.$$

- The **symmetric difference** $A \bigtriangleup B$ of A and B with respect to Ω is

$$A \bigtriangleup B = (A \setminus B) \cup (B \setminus A).$$

[1]The symbol = means equality: for the case of sets, that they contain the same elements. When you write an equality, be careful to correctly communicate for what values an equality holds and how the operations on either side of the equality are defined. In this case, I am just noting two different ways of writing the same thing.

1.2.2 Infinity and assumptions

The above definitions of intersection and union of two sets are noted to extend "nicely" for intersection and union of numerous or infinitely many sets. This assumption is fundamental to set and probability theory, but it is questionable. Infinite union gives us one example: suppose somebody can enumerate an arbitrary number of random subsets of the natural numbers. Indeed, it may be impractical for this arbitrary sequence of subsets to compute its infinite union, especially since there is no rhyme or reason to the contents of these subsets. Yet we still write mathematical notation for elements which *never* appear in any of these infinitely many subsets, as if we had a means of discovering all such elements.[2] This discussion suggests that questioning fundamental assumptions or axioms will only land you in muddy territory.

1.3 Limits of sets

Consider a sequence of subsets $\{A_k : k = 1, 2, 3, \ldots\}$. Mirroring the approach of calculus or real analysis, we define a notion of limits or "approaching a value" to these subsets in the following way:

$$\begin{aligned}
\inf_{k \geq n} A_k &:= \bigcap_{k=n}^{\infty} A_k \\
\sup_{k \geq n} A_k &:= \bigcup_{k=n}^{\infty} A_k \\
\liminf_{n \to \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \\
\limsup_{n \to \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k
\end{aligned}$$

Intuitively, $\liminf$ is the set of elements, each of which has some natural number index n for which *all* the subsets $A_n, A_{n+1}, \ldots$ contain that element. $\limsup$ is the set of all elements which, for any and all indices n, exist in some subset A_k with index $k \geq n$. Note two properties:

[2] This logic resembles the sometimes criticized law of the excluded middle, according to which all propositions are either true or false. Can we really assume that everything is either in one set or in its complement, when we lack the perhaps infinite resources required to check such a notion?

1. lim inf is a stronger idea, in that $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$.[3]

2. By de Morgan's laws, $(\liminf_{n\to\infty} A_n)^c = \limsup_{n\to\infty} A_n^c$.

Finally, we define the **limit** $\lim_{n\to\infty} A_n$ as equal to $\liminf_{n\to\infty} A_n$, or $\limsup_{n\to\infty} A_n$, if

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n;$$

otherwise, the limit does not exist. Intuitively, the limit is the set of elements which keep popping up forever as you infinitely iterate through and examine the sequence $\{A_n\}$, if these elements all eventually start popping up in every single consecutive subset. For a counterexample, the following periodic sequence of subsets has no limit:

$$\{0, 1, 2\}, \{0, 1\}, \{0, 1, 2\}, \{0, 1\}, \ldots$$

By a somewhat involved proof, one can argue that each of the following seemingly intuitive results is implied by our definitions.

$$\lim_{n\to\infty} \left(\inf_{k\geq n} A_k\right) = \liminf_{n\to\infty} A_n$$
$$\lim_{n\to\infty} \left(\sup_{k\geq n} A_k\right) = \limsup_{n\to\infty} A_n$$

1.4 Monotone sequences

A sequence of sets $\{A_n\}$ is **monotone non-decreasing**, or $A_n \nearrow$, if $A_1 \subset A_2 \subset \cdots$. A sequence is **monotone non-increasing**, or $A_n \searrow$, if $A_1 \supset A_2 \supset \cdots$.[4]

The following two facts hold for monotone sequences of sets. The proof of the first shows that $\limsup_{n\to\infty} A_n \subset \liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$; similar logic is used for the second.

1. If $\{A_n\}$ is monotone and $A_n \nearrow$, then $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n$.

2. If $\{A_n\}$ is monotone and $A_n \searrow$, then $\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n$.

It can be shown from these findings that, in generality,

$$\liminf_{n\to\infty} B_n = \lim_{n\to\infty} \left(\inf_{k\geq n} B_k\right)$$

[3] $\subset$ is a similar operation to $=$, in that they both relate two sets. Writing $A \subset B$ means that every element in A is also in B, i.e. B contains A and possibly more elements.

[4] $A \supset B$ when $B \subset A$.

and

$$\limsup_{n\to\infty} B_n = \lim_{n\to\infty}\left(\sup_{k\geq n} B_k\right).$$

1.5 Closure

A set C of subsets is said to be **closed** under a set operation if performing that operation on any subset in C returns another subset in C. This concept will be used in our conventionally accepted formal definition of a probability space. The intersection of closed sets C_t is also closed, but not necessarily the union!

A **field**, or **algebra** in the context of set theory, is a non-empty set of subsets which is closed under finite union, finite intersection, and complements. A **σ-field** or **σ-algebra** is the same idea, only stronger: it is also closed under *infinite* union and intersection. A minimal set of conditions for $\mathcal{A}$ to be a field (σ-field) is that $\Omega \in \mathcal{A}$, $\mathcal{A}$ is closed under complements, and $\mathcal{A}$ is closed under finite (infinite) union. (If these conditions are satisfied, de Morgan's laws will then imply that $\mathcal{A}$ is closed under finite (infinite) intersection.) The intersections of fields and σ-fields are closed.

1.6 A σ-field generated by C

$\sigma(C)$ is defined as the **minimal σ-field** over C: a σ-field containing C and which is also a subset of any other σ-field containing C. For a set C of subsets of Ω, some abstract math can show that the *intersection* of all σ-fields $\mathcal{B}$ containing C is $\sigma(C)$. Naturally, this conclusion is quite abstract; in the words of Sidney Resnick, "explicit construction [of such a σ-algebra] is usually hopeless."

1.7 Borel sets on the real line

The **Borel subsets** of the real numbers are the minimal σ-field containing C, or $\sigma(C)$, with

$$C = \{(a, b], -\infty \leq a \leq b < \infty\}.$$

According to some tedious proofs, we can equivalently use open or closed endpoints for either the left or the right bounds above (a or b), as long as the corresponding (left or right) ∞ term is the opposite: e.g., $\leq$ for (and $<$ for [. Any such way, we'll be dealing with the same σ-algebra.

Suppose that $\Omega_0 \subset \Omega$. If $\mathcal{B}$ is a σ-field over Ω, then $\mathcal{B}$, the σ-field containing nothing but the intersection of any element in $\mathcal{B}$ and Ω_0, is a σ-field over Ω_0.

Furthermore, consider $\mathcal{C}$ as a set of subsets of Ω, and $\mathcal{B} = \sigma(\mathcal{C})$. Defining $\mathcal{C}_0$ by "intersecting" (roughly speaking) $\mathcal{C}$ with Ω_0, one can prove that $\sigma(\mathcal{C}_0)$ is also equivalent to $\sigma(\mathcal{C})$ "intersected" with Ω_0. The proofs of these facts are somewhat involved.

2 Probability spaces

2.1 Definitions and properties

A **probability space** is a triple $(\Omega, \mathcal{B}, P)$, for which

- Ω is a set containing exactly the presumed possible outcomes of some experiment.
- $\mathcal{B}$ is a σ-algebra of subsets of Ω. These subsets are called **events**.
- P is a **probability measure**: a function with domain $\mathcal{B}$ and range $[0, 1]$, such that
  1. $P(A) \geq 0$ for all events $A \in \mathcal{B}$.
  2. P is σ-additive: If $\{A_n\}$ is a set of *disjoint*[5] events, then
  $$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$
  Because the infinite union must yield a valid event in $\mathcal{B}$ and P is defined over $\mathcal{B}$, the infinite sum must converge to the same value.
  3. $P(\Omega) = 1$.

This definition has various implications.

1. Since $A \cup A^c = \Omega$ and A and A^c are disjoint, so $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$. Note that I have just used the axiom of substitution, which allows us to freely interchange Ω and $A \cup A^c$, which are the same set, and write expressions like $P(\Omega) = P(A \cup A^c)$.
2. From property 1, we have $P(\emptyset) = 0$.

[5]Disjoint: sharing no element.

3. $P(A \cup B) = P(A) + P(B) - P(AB)$, by noting $A \cup B = AB^c \cup B$ and $A = AB \cup AB^c$ and using the σ-additive property of P.

4. The **inclusion-exclusion** formula, provable by induction, generalizes the previous point:

$$P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} P(A_j) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) - \cdots + (-1)^{n+1} P(A_1 A_2 \cdots A_n)$$

5. The **monotonicity property**: If $A \subset B$, then $P(A) \leq P(B)$ since $P(B) = P(A) + P(B \setminus A)$.

6. **Subadditivity**: By reconsidering $\cup_{n=1}^{\infty} A_n$ as the union of $A_1, A_1^c A_2, A_1^c A_2^c A_3, \ldots,$ one can obtain through σ-additivity and monotonicity that

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

7. **Continuity**: If $A_n \nearrow A$ ($A_1 \subset A_2 \subset \cdots$ and A is the limit), then $P(A_n) \nearrow P(A)$, in that the sequence whose nth element is $P(A_n)$ is non-decreasing and has limit $P(A)$. Likewise, if $A_n \searrow A$, then $P(A_n) \searrow P(A)$ (the same definition, but $P(A_n)$ is non-increasing). The proof is by construction of disjoint sets $B_1 = A_1$ and $B_i = A_i \setminus A_{i-1}$ for $i > 1$.

8. The **Fatou lemma** combines these previous properties to observe that for $A_n \in \mathcal{B}, n \geq 1$,

$$P(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} P(A_n) \leq \limsup_{n\to\infty} P(A_n) \leq P(\limsup_{n\to\infty} A_n).$$

   (Above, $\liminf_{n\to\infty} P(A_n)$ is defined as $\lim_{n\to\infty} \inf_{k\geq n} P(A_n)$, and likewise for $\limsup_{n\to\infty} P(A_n)$. Notably, we assume that these limits exist, although finding them may be hopelessly difficult.)

   An implication is that if $\lim_{n\to\infty} A_n = A$, or in less formal notation, $A_n \to A$, then $\lim_{n\to\infty} P(A_n) = P(A)$, i.e. $P(A_n) \to P(A)$.

Distribution functions

If $\Omega = \mathbb{R}$, $\mathcal{B}$ is a σ-algebra containing $(-\infty, x]$ for all $x \in \mathbb{R}$, and P is a probability measure over $\mathcal{B}$, then a **(probability) distribution function** is $F : \mathbb{R} \to [0, 1]$, defined so that $F(x) = P((-\infty, x])$ for any $x \in \mathbb{R}$. Under these assumptions,

1. F is right continuous, in that for all $x \in \mathbb{R}$ $\lim_{y \to x^+} F(y) = F(x)$,
2. F is monotone non-decreasing, and
3. F has limits of 1 at ∞ and 0 at $-\infty$.

In practice, we often know F and want to go the other direction to obtain P; more on this later.

2.2 More on closure