Wasserman: All of Statistics

Kenny Workman

March 15, 2025

1 Probability

1.1 Basics

Definition 1.1. The **sample space** Ω is the set of outcomes from an experiment. Each point is denoted ω and subsets, eg. $A \subset \Omega$ are called **events**.

Definition 1.2 (Axioms of Probability)**.** A function $\mathbb{P} : \Omega \to \mathbb{R}$ that assigns a real number to each event $A \subset \Omega$ is called a **probability function** or **probability measure** if it satsifies these three axioms:

1. **Non-negativity**. $\mathbb{P}(A) \geq 0$ for every event A
2. **Normalization**. $\mathbb{P}(\Omega) = 1$.
3. **Additivity**. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if $A \cap B = \emptyset$.

It is incredible, and not obvious, that much of probability is built up from these only these three axioms

Example 1.3. It's actually tricky to show $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ with these three facts:

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(AB^c \cap AB \cap A^cB) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^cB) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^cB) + P(AB) - P(AB) \\
&= \mathbb{P}(AB^c \cup AB) + \mathbb{P}(A^cB \cup AB) - P(AB) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - P(AB)
\end{aligned}$$

Another simple idea is that events that are identical at the limit should have identical probabilities.

Theorem 1.4 (Continuity of Events)**.** *If* $A_n \to A$ *then* $\mathbb{P}(A_n) \to \mathbb{P}(A)$.

Proof. Let A_n be monotone increasing: $A_1 \subset A_2 \subset \dots$. Let $A = \lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$.

Construct disjoint sets B_i from each A_i where $B_1 = A_1$ and $B_n = \{\omega \in \Omega : \omega \in A_n, \omega \notin \bigcup_{i=1}^{i-1} A_i\}$. It will be shown that (1) each pair of B_i are disjoint, (2) $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$ and (3) $A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ (Exercise 1.1).

From Axiom 3: $\mathbb{P}(A_n) = \mathbb{P}(\bigcup\limits_{i=1}^{n} A_i) = \mathbb{P}(\bigcup\limits_{i=1}^{n} B_i) = \sum\limits_{i=1}^{n} \mathbb{P}(B_i)$.

Then $\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \sum\limits^{n} \mathbb{P}(B_n) = \sum\limits^{\infty} \mathbb{P}(B_n) = \mathbb{P}(\bigcup\limits^{\infty} B_n) = \mathbb{P}(A)$

□

Definition 1.5 (Conditional Probability)**.** If $\mathbb{P}(B) > 0$, then the probability of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

Exercise 1.1. Fill in the details for Theorem 1.2 and extend to the case where A_n is monotone decreasing.

Proof. For any pair B_{n+1} and B_n, because $B_n \subset A_n$ and $B_{n+1} \cap A_n = \emptyset$, it follows that $B_{n+1} \cap B_n = \emptyset$.
Let $\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i$. Then $\bigcup_{i=1}^{n+1} B_i = (A_{n+1} \setminus \bigcup_{i=1}^n A_i) \bigcup (\bigcup_{i=1}^n A_i) = \bigcup_{i=1}^{n+1} A_i$.
For the monotone decreasing case, let A_n be a sequence where $A_1 \supset A_2 \supset A_3 \dots$.
Observe $A_1^c \subset A_2^c \dots$ and $\lim_{n\to\infty} A_n = \Omega \setminus \bigcup^\infty A_i^c$. Construct disjoint B_n^c from A^c in the same way.
Then $\lim_{n\to\infty} \mathbb{P}(A_n) = 1 - \sum^{\infty} \mathbb{P}(B_i^c) = 1 - \mathbb{P}(A^c) = \mathbb{P}(A)$ □

Exercise 1.3. Let Ω be a sample space and $A_1, A_2, \dots$ be events. Define $B_n = \cup_{i=n}^\infty A_i$ and $C_n = \cap_{i=n}^\infty A_i$.

(a) Show $B_1 \supset B_2 \supset B_3 \dots$ and $C_1 \subset C_2 \subset C_3 \dots$

(b) Show $\omega \in \cap_{n=1}^\infty B_n$ iff ω is in an infinite number of the events

(c) Show $\omega \in \cup n = 1^\infty C_n$ iff ω belongs to all of the events, except possibly a finite number of those events.

Proof. (a) Certainly $\cup_{i=1}^\infty A_i \supset \cup_{i=2}^\infty A_i \dots$ and $\cap_{i=1}^\infty A_i \subset \cap_{i=2}^\infty \dots$.

(b) Forward. Assume $\omega \in \cap_{n=1}^\infty B_n$. If ω does not belong to an infinite number of events A_i, there exists some index j past which $\omega \notin B_j$. Then certainly $\omega \notin \cap_{n=1}^\infty B_n$. Reverse. ω belonging to infinite events means there cannot exist such a j described previously so $\omega \in B_n$ for all n. Indeed $\omega \in \cap_{n=1}^\infty B_n$

(c) Forward. Assume $\omega \in \cup_{n=1}^\infty C_n$. Then $\omega \in C_j = \cap_{i=j}^\infty A_i$ for some j. This is another way of saying ω is in every single event except for perhaps a finite number in $A_{i<j}$. Reverse. Let j be the index of the largest event that ω is not in. Then $\omega \in C_{n>j}$ and certainly $\omega \in \cup^\infty C_n$.

□

Note. The key idea above is this notion of "infinitely often" (i.o.) and "all but finitely often" (eventually) which are two distinct structures of infinite occurence in sequences. Consider an ω that exists in every other event (eg. just the odd indices) for infinite events and revisit its inclusion in $\cap^\infty B_i$ and $\cup^\infty C_i$.

Note. $\lim \cap \cup A_n$ is also referred to as the limit infimum of A_n. Similarly, $\lim \cup \cap A_n$ is referred to as the limit supremum of A_n.

Exercise 1.7. Let $\mathbb{P}(\bigcup^n A_i) \le \sum^n \mathbb{P}(A_i)$. Then $\mathbb{P}(A_{n+1} \cup (\bigcup^n A_i)) \le \mathbb{P}(A_{n+1}) + (\sum^n \mathbb{P}(A_i)) - \mathbb{P}(A_{n+1} \cap (\bigcup^n A_i)) \le \sum^{n+1} \mathbb{P}(A_i)$

Note. Expand a bit on the Boole inequality.

Exercise 1.9. For fixed B s.t. $\mathbb{P}(B) > 0$, show $\mathbb{P}(. \mid B)$ satisfies the three axioms of probability.

Proof.

- **Non-negativity**. If $\mathbb{P}(B) > 0$ and $\mathbb{P}(AB) > 0$ for any $A \subset \Omega$, certainly $\frac{\mathbb{P}(AB)}{\mathbb{P}(B)} > 0$.
- **Normalization**. $\frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$
- **Additivity**. Let $AB \cap CB = \emptyset$, then $\mathbb{P}(AB \cap CB) = \mathbb{P}(AB) + \mathbb{P}(CB)$. Indeed $\frac{\mathbb{P}(AB \cap CB)}{B} = \frac{\mathbb{P}(AB)}{B} + \frac{\mathbb{P}(CB)}{B}$

□

Exercise 1.11. Suppose A and B are independent events. Show that A^c and B^c are also independent.

Proof. We are given $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$. Then $\mathbb{P}(A^c)\mathbb{P}(B^c) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) = 1 - \mathbb{P}(A \cup B) = \mathbb{P}(A^c B^c)$. The second to last equality uses independence of $P(AB)$. The last equality uses the property of set complements $P(A \cup B) = P(A^c \cap B^c)$. □

Exercise 1.13. Suppose a fair coin is tossed repeatedly until heads and tails is each encoutered exactly once. Describe Ω and compute the probablity exactly three tosses are needed.

Proof.

- The sample space is the set of binary strings with exactly one 0 and 1. For strings of length greater than 2, these are repeated strings of 0 or 1 capped with a 1 or 0 respectively.

- By independence, each n-string has an identical probability $\frac{1}{2}^n$. There are two such 3-strings: 001 and 110. Using additivity, $\mathbb{P}(3 \text{ tosses}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$

$\square$

Exercise 1.15. The probability a child has blue eyes is $\frac{1}{4}$. Assume independence between children. Consider a family with 3 children.

- If it is known that at least one of the children have blue eyes, what is the probablity that at least two of the children have blue eyes?
- If it is know that the youngest child has blue eyes, what is the probability that at least two of the children have blue eyes?

Proof.

- Straightforward conditional probability. Let A be the event where at least one child has blue eyes and B be the event where at least two children have blue eyes. Consider first, $\mathbb{P}(A) = 1 - \mathbb{P}(\text{no child has blue eyes}) = 1 - \frac{27}{64} = \frac{37}{64}$. Compute $\mathbb{P}(A \cap B)$ by enumerating events 101, 111, 110 and using additivity: $2 \cdot \frac{1}{4}^2 \cdot \frac{3}{4} + \frac{1}{4}^3 = \frac{10}{64}$. Then $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{10}{64} \cdot \frac{64}{37} = \frac{10}{37}$
- Similar procedure. Let A be the event where the youngest child has blue eyes and B be as before. Using independence, $\mathbb{P}(A) = \frac{1}{4}$. (To see this rigorously, enumerate the sample space and see $\mathbb{P}(\Omega \mid \text{first child blue}) = 1$). Now $\mathbb{P}(B \cap A)$ describe events 110, 101, 111 only. $\frac{7}{64} \cdot \frac{4}{1} = \frac{7}{16}$.

$\square$