Probability_Basics

December 15, 2020

Table of Contents

1 Probability Basics

1.1 Some boring mathematical definitions

1.1.1 Experiment (probability)

1.1.2 Some Notation

1.1.3 Axioms of Probability

1.2 Useful Properties of Probability (Derived from Axioms)

1.2.1 The Complement Rule

1.2.1.1 Proof

1.2.2 Numeric Bound

1.2.2.1 Proof

1.2.3 Probability of Empty Set

1.2.3.1 Proof

1.2.4 Monotonicity

1.2.4.1 Proof

1.2.5 Rule of Addition

1.2.5.1 Proof

1.3 Some Useful Definitions

1.3.1 Independence

1.3.2 Mutually Exclusive

1.3.3 Conditional Probability

1.3.4 Bayes Theorem

1.3.4.1 Proof

1.4 Practical Probability Mathematics

1.4.1 Permutations

1.4.1.1 Definition

1.4.1.2 Dealing With Repeated Elements

1.4.2 Combination

1.4.3 Sums

1.4.3.1 Arithmetic Sum

1.4.3.2 Proof

1.4.4 Geometric Sum

1.4.4.1 Finite Case

1.4.4.2 Infinite Case

Author: DistilledData

Contents

1 Probability Basics **3**
- 1.1 Some boring mathematical definitions . . . 3
  - 1.1.1 Experiment (probability) . . . 3
  - 1.1.2 Some Notation . . . 3
  - 1.1.3 Axioms of Probability . . . 3
- 1.2 Useful Properties of Probability (Derived from Axioms) . . . 3
  - 1.2.1 The Complement Rule . . . 3
  - 1.2.2 Numeric Bound . . . 4
  - 1.2.3 Probability of Empty Set . . . 4
  - 1.2.4 Monotonicity . . . 4
  - 1.2.5 Rule of Addition . . . 5
- 1.3 Some Useful Definitions . . . 6
  - 1.3.1 Independence . . . 6
  - 1.3.2 Mutually Exclusive . . . 6
  - 1.3.3 Conditional Probability . . . 6
  - 1.3.4 Bayes Theorem . . . 6
- 1.4 Practical Probability Mathematics . . . 6
  - 1.4.1 Permutations . . . 6
  - 1.4.2 Combination . . . 7
  - 1.4.3 Sums . . . 7
  - 1.4.4 Geometric Sum . . . 8

1 Probability Basics

This notebook is meant to serve as a study guide for the highlights of probability theory that is used in data science. The more mathematically rigorous parts of this notebook require an appreciation of set theory.

1.1 Some boring mathematical definitions

1.1.1 Experiment (probability)

Outcome: a possible result of an experiment (e.g. flipping a coin may result in H or T) **Experiment**: A procedure with a fixed set of possible outcomes that may be repeated indefinitely. An experiment has three parts: 1. A **Sample Space** (S or Ω): set of all possible outcomes 2. A set of **events** ($\mathcal{F}$) (an **event** is a set containing zero or more outcomes) 3. Assignment of probabilites to the events (i.e. a function P mapping events to probabilities. Furthermore, we require that as the number times of an experiment tends towards infinity, the frequency of the event $\mathcal{F}$ approaches $P(\mathcal{F})$

Probability: A function P that maps each event in $\mathcal{F}$ to a real number between 0 and 1, inclusive

1.1.2 Some Notation

1. Given an event $A \in \mathcal{F}$, the complement of A, A^c, is all outcomes that are not in A (i.e. $\Omega \backslash A$, or all outcomes $B \in \otimes$ such that $B \notin A$)
2. Given two events A and B, $A \cap B$ is the set of outcomes that are in both A and B (i.e. $\forall \omega \in A \cap B$, $\omega \in A$ and $\omega \in B$
3. Given two events A and B, $A \cup B$ is the set of outcomes that are in either A or B (i.e. $\forall \omega \in A \cup B$, $\omega \in A$ or $\omega \in B$

1.1.3 Axioms of Probability

1. **Non-Negativity** $\forall E \in \mathcal{F}$, $P(E) \in \mathbb{R}$ such that $P(E) \geq 0$
2. **Unitarity** $P(\Omega) = 1$
3. **σ-additivity**

For all countable sequences of disjoint sets, E_1, E_2, $\cdots$, $P\big(\bigcup\limits_{i=1}^{\infty} E_i\big) = \sum\limits_{i=1}^{\infty} P(E_i)$

1.2 Useful Properties of Probability (Derived from Axioms)

1.2.1 The Complement Rule

$P(A^c) = P(\Omega \backslash A) = 1 - P(A)$

Proof **Lemma:** $\forall A \in \mathcal{F},\ A \cup A^c = \Omega$

If $\omega \in A \cup A^c$, then $\omega \in A$ or $\omega \in A^c$. Since $A \subseteq \Omega$ and $A^c \subseteq \Omega$, $\omega \in \Omega$

If $\omega \in \Omega$, since $A \subseteq \Omega$, we have two possibilities: 1) $\omega \in A$ 2) $\omega \notin A$

If $\omega \notin A$, then by definition $\omega \in A^c$, which means that $\omega \in A \cup A^c$ and hence $\Omega \subseteq A \cup A^c$

$\therefore A \cup A^c = \Omega$

Claim: $P(A^c) = P(\Omega \backslash A) = 1 - P(A)$

Using the fact that A^c and A are disjoint such that $A \cup A^c = \Omega$, by the σ-additivity and unitarity axioms, we have

$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) \implies P(A^c) = 1 - P(A)$

1.2.2 Numeric Bound

$\forall E \in \mathcal{F},\ 0 \leq P(E) \leq 1$

Proof $\forall E \in \mathcal{F}$, axiom 1 gives us the numberic bound $P(E) \geq 0$, which leaves us to show $P(E) \leq 1$. Using the complement rule above, we obtain

$P(E^c) = 1 - P(E) \geq 0 \implies 1 \geq P(E)$

$\therefore 0 \leq P(E) \leq 1$

1.2.3 Probability of Empty Set

$P(\emptyset) = 0$

Proof **Lemma:** $\Omega^c = \emptyset$

$\forall \omega \in \Omega^c$, by definition $\omega \notin \Omega$. However, Ω contains all possible outcomes, which is a contradiction. Hence $\Omega^c = \emptyset$

Claim: $P(\emptyset) = 0$

Since, $\emptyset$ is the complement of Ω, we can use the complement rule above and axiom 2. $P(\Omega^c) = P(\emptyset) = 1 - P(\Omega) = 1 - 1 = 0$

1.2.4 Monotonicity

$\forall A, B \in \mathcal{F}$ such that $A \subseteq B$, $P(A) \leq P(B)$

Proof **Lemma:** $\forall A, B \in \mathcal{F}$ **such that** $A \subseteq B$**,** $B = A \cup (B \backslash A)$

If $\omega \in B$, since $A \subseteq B$, either $\omega \in A$ or $\omega \notin A$. If $\omega \in A$, then $\omega \in A \cup (B \backslash A)$. If $\omega \notin A$, then $\omega \in B \backslash A$ and hence $\omega \in A \cup (B \backslash A)$. Hence, $B \subseteq A \cup (B \backslash A)$.

If $\omega \in A \cup (B \backslash A)$, then either $\omega \in A$ or $\omega \in (B \backslash A)$. If $\omega \in A$, then $\omega \in B$ since $A \subseteq B$. If $\omega \in (B \backslash A)$, then by definition $\omega \in B$. Hence, $A \cup (B \backslash A) \subseteq B$.

$\therefore B = A \cup (B \backslash A)$

Claim: $\forall A, B \in \mathcal{F}$ **such that** $A \subseteq B$**,** $P(A) \leq P(B)$

Since events A and $B \backslash A$ are disjoint where $B = A \cup (B \backslash A)$, we have $P(B) = P(A \cup (B \backslash A)) = P(A) + P(B \backslash A)$

Since we have shown above (Numberic Bound) $\forall \omega \in \mathcal{F}$, $0 \leq P(\omega) \leq 1$, we have $0 \leq P(B \backslash A) \leq 1$, and hence $P(A) = P(B) - P(B \backslash A) \leq P(B)$.

$\therefore P(A) \leq P(B)$

1.2.5 Rule of Addition

$\forall A, B \in \mathcal{F}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof **Lemma 1:** $\forall A, B \in \mathcal{F}$**,** $B \backslash A = B \backslash (A \cap B)$

If $\omega \in B \backslash A$, then by definition $\omega \in B$ and $\omega \notin A$. Since $\omega \notin A$, $\omega \notin A \cap B$ and hence $\omega \in B \backslash (A \cap B) \implies B \backslash A \subseteq B \backslash (A \cap B)$

If $\omega \in B \backslash (A \cap B)$, then by definition $\omega \in B$ and also by definition $\omega \notin (A \cap B) \implies \omega \notin A$. Thus $\omega \in B \backslash A$ and hence $B \backslash (A \cap B) \subseteq B \backslash A$

$\therefore \forall A, B \in \mathcal{F}$, $B \backslash A = B \backslash (A \cap B)$

Lemma 2: $\forall A, B \in \mathcal{F}$**,** $A \cup B = A \cup (B \backslash A)$

We will first show $A \cup B = A \cup (B \backslash A \cap B)$ and use lemma 1 to prove lemma 2.

$\forall \omega \in A \cup B$, by definition $\omega \in A$ or $\omega \in B$. If $\omega \in A$, then $\omega \in A \cup (B \backslash A \cap B)$. If $\omega \in B$, then either $\omega \in A \cap B$ and hence $\omega \in A$, or $\omega \in B \backslash A \cap B$, which means $\omega \in A \cup (B \backslash A \cap B)$. Thus, $A \cup B \subseteq A \cup (B \backslash A \cap B)$.

$\forall \omega \in A \cup (B \backslash A \cap B)$, by definition $\omega \in A$ or $\omega \in B \backslash A \cap B$. If $\omega \in A$, then $\omega \in A \cap B$. If $\omega \in B \backslash A \cap B$, then by definition $\omega \in B$ and hence $\omega \in A \cup B$. Thus, $A \cup (B \backslash A \cap B) \subseteq A \cup B$.

$\therefore A \cup B = A \cup (B \backslash A \cap B)$ and using Lemma 1 above, we have $A \cup B = A \cup (B \backslash A)$

Claim: $\forall A, B \in \mathcal{F}$**,** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

By the σ-additivity axiom on disjoint sets A and $B \backslash A$, and the lemmas above, we have $P(A \cup B) = P(A \cup (B \backslash A)) = P(A) + P(B \backslash A) = P(A) + P(B \backslash (A \cap B))$

Using the σ-addivity axiom on disjoint sets $A \cap B$ and $B \backslash (A \cap B)$, we obtain $P(B) = P\Bigg(\Big(A \cap B\Big) \cup \Big(B \backslash (A \cap B)\Big)\Bigg) = P(A \cap B) + P(B \backslash (A \cap B))$

$$\implies P(A \cup B) = P(A) + P(B \backslash (A \cup B)) = P(A) + P(B) - P(A \cap B)$$

1.3 Some Useful Definitions

1.3.1 Independence

$\forall A, B \in \mathcal{F}$, A and B are independent $\iff P(A \cap B) = P(A)P(B)$

1.3.2 Mutually Exclusive

$\forall A, B \in \mathcal{F}$, A and B are mutually exclusive $\iff P(A \cap B) = 0$

1.3.3 Conditional Probability

$\forall A, B \in \mathcal{F},\ P(A|B) = \frac{P(A \cap B)}{P(B)}$

1.3.4 Bayes Theorem

$\forall A, B \in \mathcal{F},\ P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Proof $P(A|B) = \frac{P(A \cap B)}{P(B)}$

$\frac{P(B|A)P(A)}{P(B)} = \frac{\frac{P(B \cap A)}{P(A)}P(A)}{P(B)} = \frac{P(B \cap A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$

$\therefore P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

1.4 Practical Probability Mathematics

1.4.1 Permutations

Definition A set or collection of items where order matters. For instance, each sequence of characters that form a license plate is a distinct permutation. Another example is choosing a card from a deck of cards without replacement. The sequence of cards that is chosen form a permutation.

Given a set of n items from which we want to choose k items without replacement in an order-dependent fashion, the number of possibilities is given by

$$P(n, k) = \frac{n!}{(n-k)!}$$

Dealing With Repeated Elements Suppose an item occurs r times in our set. The number of permutations then is given by

$$P(n,k) = \frac{n!}{r!(n-k)!}$$

Example: How many different permutations are there of the letters in the word "gooogle"?

Answer: The word "gooogle" has 7 letters. Since we have three 'o's and two 'g's, the total number of permutations is given by $\frac{7!}{3!2!}$

1.4.2 Combination

A combination is an order-independent set or collection of items. For instance, if you are choosing three people for a team out of ten people, the order that you choose the people does not matter, just the total number of distinct teams.

The number of combinations for choosing r items out of n possibilities is given by:

$$C(n,k) = \frac{n!}{k!(n-k)!}$$

This corresponds to a permutation where an element has k repeats. This makes intuitive sense because for every distinct permutation of k items (distinct here means that the count of each individual item type is unique since order does not matter), there are k! arrangements of those items that we consider "repeats".

1.4.3 Sums

Arithmetic Sum An arithmetic sequence is a list of integers $a_1, a_2, \cdots, a_n$ such that $\forall i \in \{2, \cdots, n\}$ $a_i - a_{i-1} = d$ for some $d \in \mathbb{N}$ where $d > 0$

If we sum this arithmetic sequence with n terms, initial element a_0, last element a_n, and difference d between terms, we generate an arithmetic sum whose formula is shown below:

$$S = \frac{(a_0 + a_n)n}{2}$$

Proof **Lemma:** $\sum\limits_{i=1}^{n}(i-1) = \frac{n(n-1)}{2}$

If n is even, then we have

$\sum\limits_{i=1}^{n}(i-1) = \left(\sum\limits_{i=1}^{n} i\right) - n$

$\sum\limits_{i=1}^{n} i = \sum\limits_{i=1}^{n/2} i + \sum\limits_{i=(n/2)+1}^{n} i = \sum\limits_{i=1}^{n/2} i + \sum\limits_{i=1}^{n/2} (\frac{n}{2} + i)$

$\sum_{i=1}^{n/2} i$ corresponds to summing the numbers 1 to $\frac{n}{2}$. $\forall w \in \{1, \cdots, \frac{n}{2}\} \exists 1 \leq k \leq \frac{n}{2}$ such that $w + k = \frac{n}{2} + 1 \implies w = \frac{n}{2} - k + 1$, which allows us to rewrite our summation as $\sum_{i=1}^{n/2} (\frac{n}{2} - i + 1)$

$\implies \sum_{i=1}^{n} i = \sum_{i=1}^{n/2} (\frac{n}{2} - i + 1) + \sum_{i=1}^{n/2} (\frac{n}{2} + i) = \sum_{i=1}^{n/2} (n+1) = \frac{n}{2}(n+1)$

$\implies \sum_{i=1}^{n} (i-1) = \left(\sum_{i=1}^{n} i \right) - n = \frac{n^2}{2} + \frac{n}{2} - n = \frac{n^2}{2} - \frac{n}{2} = \frac{n(n-1)}{2}$

Now suppose that n is odd and hence $n-1$ is even. Using the fact that $\sum_{i=1}^{n} (i-1) = \sum_{i=2}^{n} (i-1) = \sum_{i=1}^{n-1} i$ and using similar logic to the case when n is even, we obtain

$\sum_{i=1}^{n-1} i = \sum_{i=1}^{(n-1)/2} (\frac{n-1}{2} - i + 1) + \sum_{i=1}^{(n-1)/2} (\frac{n-1}{2} + i) = \sum_{i=1}^{(n-1)/2} n = \frac{n(n-1)}{2}$

$\therefore \sum_{i=1}^{n} (i-1) = \frac{n(n-1)}{2}$

Claim: $S = \frac{(a_0 + a_n)n}{2}$

$S = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_1 + (i-1) \cdot d = n \cdot a_1 + d \sum_{i=1}^{n} (i-1) = n \cdot a_1 + d \frac{n(n-1)}{2} = \frac{n}{2}\big(2a_1 + d(n-1)\big)$

Using the fact that $a_n = a_1 + (n-1)d$, we obtain

$S = \frac{n}{2}\big(2a_1 + d(n-1)\big) = \frac{n}{2}\big(a_1 + a_n\big)$

1.4.4 Geometric Sum

Finite Case Given a non-zero number $a_0 \in \mathbb{R}$ and a rate $r \in \mathbb{R}$ such that $r \neq 1$, a finite geometric series with n terms is a sequence of numbers $a_0, r \cdot a_0, \ldots, r^{n-1} \cdot a_0$.

The sum of the terms in a geometric series is called a geometric sum. The formula for a geometric sum is given by

$$S_n = \frac{(1 - r^n)a_0}{1 - r}$$

Proof Multiplying both sides of the summation by r and subtracting the two equations, we obtain

$S_n = a_0 + r \cdot a_0 + \cdots + r^{n-1} \cdot a_0$

$r \cdot S_n = r \cdot (a_0 + r \cdot a_0 + \cdots + r^{n-1} \cdot a_0) = r \cdot a_0 + r^2 \cdot a_0 + \cdots + r^n \cdot a_0$

$\implies S_n - r \cdot S_n = (a_0 + r \cdot a_0 + \cdots + r^{n-1} \cdot a_0) - (r \cdot a_0 + r^2 \cdot a_0 + \cdots + r^n \cdot a_0)$

$(1 - r)S_n = a_0 - r^n \cdot a_0 = (1 - r^n)a_0$

$\therefore S_n = \frac{(1-r^n)a_0}{1-r}$

Infinite Case In order to for a sum of an infinite geometric sum to converge to a value, r must satisfy $0 \leq r < 1$. Given such an r, the value for the sum becomes

$$S_n = \frac{a_0}{1-r}$$

Proof $\lim\limits_{n\to\infty} S_n = \lim\limits_{n\to\infty} \frac{(1-r^n)a_0}{1-r} = \lim\limits_{n\to\infty} \frac{a_0}{1-r} + \lim\limits_{n\to\infty} \frac{r^n a_0}{1-r}$

Since $0 \leq r < 1$, $\lim\limits_{n\to\infty} r^n = 0$ and hence $\lim\limits_{n\to\infty} \frac{r^n a_0}{1-r} = 0$

We also know that $\frac{a_0}{1-r}$ has no dependence on n and hence $\lim\limits_{n\to\infty} \frac{a_0}{1-r} = \frac{a_0}{1-r}$

$\implies S_\infty = \lim\limits_{n\to\infty} \frac{a_0}{1-r} + \lim\limits_{n\to\infty} \frac{r^n a_0}{1-r} = \frac{a_0}{1-r}$