PROBABILITY - INTRODUCTION

Jerzy Szulga

Department of Mathematics and Statistics
Auburn University

MATH 5670-6670 FALL 2019

PROBABILITY I

August 19, 2019

Outline

Probability as proportion

A scientific approach to probability arose gradually.

The probability of an event composed of finitely many equally likely possible outcomes:

$$\frac{\textbf{number of outcomes}}{\textbf{total number of outcomes}}$$

(From ancient times, to Gerolamo Cardano, to Blaise Pascal, to Rev. Thomas Bayes, to Marquis de Laplace, 17th-19th century).

Laplace's wording

The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

Gibbs' energy-related probability

A particle in a "canonical ensemble" with the Helmholtz free energy A stays in a state $j = 1, 2, 3, \ldots$ with energy E_j with probability P_j proportional to $Q_j = e^{-k_B E_j/T}$, where k_B is the Boltzmann constant and T is the absolute temperature. The greater the energy, the smaller the probability - but exponentially, not proportionally.

The fundamental free energy equation yields the probabilities:

$$Q = e^{-k_B A/T} = \sum_j Q_j \quad \Rightarrow \quad p_j = \frac{Q_j}{Q}$$

(Josiah Willard Gibbs' *Statistical Mechanics*, end of 19th century)

Bernoulli Process

A.k.a. **Bernoulli trials**.

An experiment or observation, a.k.a. a **trial** -

- is repeated independently and identically,
- has two possible random outcomes 1 (yes) or 0 (no).

A record is a finitary (i.e., finite with varying length) 01-sequence.

The probability of the event is the **relative frequency**

$$\frac{\text{number of 1s}}{\text{total number of trials}}$$

Two modern mainstream approaches

- **frequential**:

  practical probability as the frequency of an event in Bernoulli trials

- **Bayesian**:

  ideal or **theoretical** probability as a measurement of an event, modeled as a subset of a sample space, subject to simple rules.

Other approaches?

- logic of uncertainty
- information and entropy
- computational complexity
- fuzzy sets theory
- quantum probability and the collapsed wave function (Schrödinger's cat etc.)
- ....

“Psychological probability”

People use it every day as an expression of degree of a belief. e.g.,

- *there is a 50% chance that I'll pass the exam*
- *there is a 1% chance that the Moon is hollow*
- *there is a 99% chance that the Sun will rise tomorrow*
- etc. (more examples in Ross)

Ross argues (Sect. 2.7) that such “measure of belief” basically satisfies axioms of probability although sometimes it doesn't.

What are chances that he is right?

Distinctive terminology

The basis: set theory and formal logic.

The language: specific to probability.

- A "**set**" is not defined but is expected to be self explanatory. Now it is called an **event**, e.g., "**an event has occurred**".
- A "mother set" of possible outcomes, now: a **sample space**.
- A set being a **subset** of another, $A \subset B$, now: **the event** A **implies the event** B.

Venn diagrams

- Draw the "mother set" as a rectangle;
- Draw sets as ovals or other figures inside the rectangle;
- Illustrate the basic set operations -

  the union, the intersection, the complement, etc.

Warning:

Venn diagrams work great with 2 sets, fine with 3 sets, but...

More sets = trouble (tedious, unreliable, mistake prone).

- An element a **belongs to** a set A, $a \in A$,

  now: **an event A includes an outcome a**.

- the **union** $A \cup B$: alternative A **or** B,

  now: **one of two events has occurred**

- the intersection $AB = A \cap B$: conjunction A **and** B,

  now: **both events have occurred**

- the complement A' or A^c: negation **not** A,

  now: **the opposed event has occurred**

- the empty set $\emptyset$,

  now: **impossible event**

- disjoint sets $A \cap B = \emptyset$,

  now: **the events are mutually exclusive**.

Reading symbols and matching logic

- $A \cup B$: **alternative**

  read: "*A or B*", "*A union B*"

- $AB = A \cap B$: **conjunction**

  read: "*A and B*", "*A intersection B*"

- A': **negation**

  read: "*not A*", "*A complement*", "*complement of A*"

- $A \subset B$: **implication**

  read: "*A implies B*", "A is a subset of B"

Beware of the Tower of Babel

The symbolic notation is one and only one, rarely changes, and is used in the same form all over the planet Earth and beyond.

In contrast, verbal expressions change from one language to another, one region to another, between groups and individuals.

E.g., the crisp and clear $A \subset B$ can be read (in just one language) as:

- an event A entails the occurrence of the event B,
- if A happens, so will B,
- if A was observed, then necessarily B must follow,
- B is a consequence of A, etc., etc., etc., ...

The modern approach: Three Axioms

1. A sample space S, a set of possible outcomes.
2. A specified family of events (subsets) of S:
   - containing the whole space S (the **certain event**);
   - closed under countable unions and complements $A' = S \setminus A$
3. A measurement $\mathrm{P}(A) \in [0,1]$ of each event A such that:
   - $0 \le \mathrm{P}(A) \le 1 = \mathrm{P}(S)$,
   - For disjoint A_i, $\mathrm{P}(A_1 \cup A_2 \cup \cdots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots$.

(known $<$ 1920s, adopted by Kolmogorov in 1930's, in use $\geq$ 1950s)

Original Kolmogorov's axioms

First, the family of events was closed under finite unions.

Secondly, the countable additivity axiom was divided into two axioms.

(We will write $A \sqcup B$ when $A \cap B = \emptyset$.)

- **Finite additivity**: $\mathrm{P}(A \sqcup B) = \mathrm{P}(A) + \mathrm{P}(B)$;
- **Continuity**: for a sequence A_n of events descending to the empty set, i.e. $A_{n+1} \subset A_n$ for every n, and $\bigcap\limits_n \downarrow A_n = \emptyset$,

$$\lim_n \downarrow \mathrm{P}(A_n) = 0.$$

New set operations

- the **set difference**

  $A \setminus B = A \cap B'$, the part of A not in B;

- the **symmetric difference**

  $A \oplus B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$;

  "either or" - *exclusive alternative* in logic

  "or" - *inclusive alternative* in logic

Rules and Laws

- **Distributive laws**:

  $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ and

  $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

- **de Morgan's laws**:

  $(A \cup B)' = A' \cap B'$ and $(A \cap B)' = A' \cup B'$

- **Associative laws**:

  $(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C,$

  $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C = ABC,$

  $(A \oplus B) \oplus C = A \oplus (B \oplus C) = A \oplus B \oplus C$

Typical quiz or exam questions

True or false?

$(A \setminus B) \setminus C = A \setminus (B \setminus C)$?

$(A \oplus B) \cup C = (A \cup C) \oplus (B \cup C)$?

$(A \cup B) \oplus C = (A \oplus C) \cup (B \oplus C)$?

$(A \oplus B) \cap C = (A \cap C) \oplus (B \cap C)$?

The first identity - an alleged associative law.

The next three identities - alleged distributive laws.

Algebra 01

A set identity $A = B$ formally is proved by the logical equivalence:

$$x \in A \quad \Leftrightarrow \quad x \in B$$

Let us code the sentences as Boolean (true-false, 1-0) binary numbers

$$A \leftrightarrow a = \mathbb{1}_{\{x \in A\}}, \quad B \leftrightarrow b = \mathbb{1}_{\{x \in B\}}, \text{ and so on.}$$

This yields the codes for operations:

$$AB \leftrightarrow ab, \qquad A' \leftrightarrow 1 - a, \qquad A \cup B \leftrightarrow a + b - ab$$

Algebra 01, continued

Algebra 01 is a sub-category of the ordinary Algebra that yields sometimes surprising formulas, e.g.

$$a^2 = a, \quad \text{hence} \quad |a-b| = (a-b)^2 = a^2 + b^2 - 2ab = a + b - 2ab.$$

Also, $(1-a-b)(a-b) = 0$, etc.

The difference and symmetric difference of sets:

$$A \setminus B \quad \leftrightarrow \quad a(1-b), \qquad A \oplus B \quad \leftrightarrow \quad |a-b| = a + b - 2ab.$$

Algebra 01, continued

Example 1. The symmetric difference is associative,

$$(A \oplus B) \oplus C = A \oplus (B \oplus C).$$

Proof.

$$\begin{aligned} \text{LHS:} \quad & (a+b-2ab)+c-2(a+b-2ab)c \\ & = a+b-2ab+c-2ac-2bc+4abc \end{aligned}$$

$$\begin{aligned} \text{RHS:} \quad & a+(b+c-2bc)-2a(b+c-2bc) \\ & = a+b+c-2bc-2ab-2ac+4abc \end{aligned}$$

Algebra 01, continued

Example 2. Proving the first de Morgan's law $(A \cup B)' = A' \cap B'$

$$\begin{aligned} \text{LHS:} \quad & 1 - (a + b - ab) = 1 - a - b + ab \\ \text{RHS:} \quad & (1 - a)(1 - b) = 1 - a - b + ab \end{aligned}$$

Proving the second de Morgan's law $(A \cap B)' = A' \cup B'$:

$$\begin{aligned} \text{LHS:} \quad & 1 - ab \\ \text{RHS:} \quad & (1 - a) + (1 - b) - (1 - a)(1 - b) \\ = \quad & 2 - a - b - 1 + a + b - ab = 1 - ab \end{aligned}$$

Intuition

- Probability is a nonnegative **additive** quantity:

  The whole equals the sum of parts,

  akin to the count, length, area, volume, mass, energy, etc..
- Probability is **normalized**:

  The whole is assigned the value 1,
- Probability is a **pure scalar** - it has no physical unit.

Hence, the normalized area or mass of a lamina in the plane is a popular, intuitive, and convenient model for probability.

Derived "Rules" a.k.a. "Theorems"

- $\mathrm{P}(A') = 1 - \mathrm{P}(A)$.

  Draw a Venn diagram.

- $\mathrm{P}(\emptyset) = 0$.

  Denote $x = \mathrm{P}(\emptyset) = \mathrm{P}(\emptyset \cup \emptyset) = \mathrm{P}(\emptyset) + \mathrm{P}(\emptyset) = 2x$.

  Solve the equation $x = 2x$. Indeed, $x = 0$.

- $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$.

  Indeed, draw a Venn diagram and think: "area".

- $\mathrm{P}(A \cup B \cup C) =$

  $\mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(AB) - \mathrm{P}(BC) - P(AC) + \mathrm{P}(ABC)$.

  S.a.a.

Measure "Theory"

The union $A \cup B$ of two disjoint sets A and b may be written as

$$A \sqcup B \quad \text{or} \quad A \dot{\cup} B.$$

A **measure** is a nonnegative quantity Q, assigned to sets such that

- $Q(A_1 \sqcup A_2 \sqcup \cdots) = Q(A_1) + Q(A_2) + \cdots$
- $Q(\emptyset) = 0.$

We may consider measures with infinite values.

A finite measure satisfies the second condition automatically.

Probability is a normalized finite measure, $\mathrm{P}(\Omega) = 1$.

Why "measure" is avoided in spite of its simplicity

First, its domain - a certain family of sets, must be carefully examined. We see that the domain should contain $\emptyset$, the whole set Ω, and be closed under countable disjoint unions.

However, to avoid trivialities, some other set operations must be performable on the domain; perhaps, the complement A', perhaps the set difference $A \setminus B$, or so.

This is how the Measure Theory starts but in fact it originated at the end of the 19$^{\text{th}}$ century with a simple question:

Which planar sets do admit area?