NYUAD

Kundyz Muktar

Formalization of Probability Space in Isabelle, FOI

RESEARCH SEMINAR IN MATHEMATICS, 2023

Goals:

- Implementing Probability Theory in the automated theorem prover Isabelle.
- Developing a system to check the consistency of assumptions in First-Order logic.

First-Order Logic VS Second-Order Logic

FOL

Most of the probability theory can be formulated in FOL, which deals with the quantification of variables over individuals in a domain.
First-order logic allows for the definition of probability functions over sets of individuals, events, or propositions, and the calculation of probabilities using axioms and rules of inference.

SOL

SOL, on the other hand, extends FOL by allowing quantification over sets or properties of individuals. It enables the expression of more complex statements involving properties, relations, or functions over sets of individuals. It can be used to express statements about the properties of probability distributions, such as the existence of moments or the convergence of sequences of random variables.

Compactness Theorem

Theorem (Compactness)

For every set Γ of formulas it holds:

$$\Gamma \text{ is consistent} \iff \text{every finite subset } \Gamma_0 \subseteq \Gamma \text{ is consistent}$$

Proof of $\Rightarrow$

Every model for Γ also is a model for all subsets $\Gamma_0 \subseteq \Gamma$.

Proof of $\Leftarrow$ (by transposition)

- Suppose that Γ is **not** consistent. Then it follows: $\Gamma \vDash \bot$.
- Then by the completeness theorem: $\Gamma \vdash \bot$.
- Hence there is a derivation $\mathcal{D}$ of $\bot$ from premises in Γ.
- As $\mathcal{D}$ is finite, it can only use finitely many premises $\Gamma_0 \subseteq \Gamma$.
- Then $\mathcal{D}$ also shows $\Gamma_0 \vdash \bot$.
- By the soundness theorem it follows: $\Gamma_0 \vDash \bot$.
- Hence the finite subset Γ_0 of Γ is **not** consistent.

Why First-Order?

First-order logic is complete in the sense that for any first-order theory T and any first-order sentence φ, either T ⊢ φ (i.e., φ is provable from T) or T ⊢ ¬φ (i.e., the negation of φ is provable from T). This means that first-order logic can capture all the valid inferences that can be made from a given first-order theory, and that there are no true statements that cannot be proven in first-order logic.

In contrast, second-order logic is incomplete in the sense that there are true statements that cannot be proven in second-order logic. This is because the semantics of second-order logic allows for non-standard interpretations of second-order quantifiers, which cannot be captured by any finite axiomatization.

https://en.wikipedia.org/wiki/G%C3%B6del%27s_completeness_theorem#:~:text=Ackermann%20proof%20system.-,G%C3%B6del's%20original%20formulation,all%20the%20logically%20valid%20formulae.

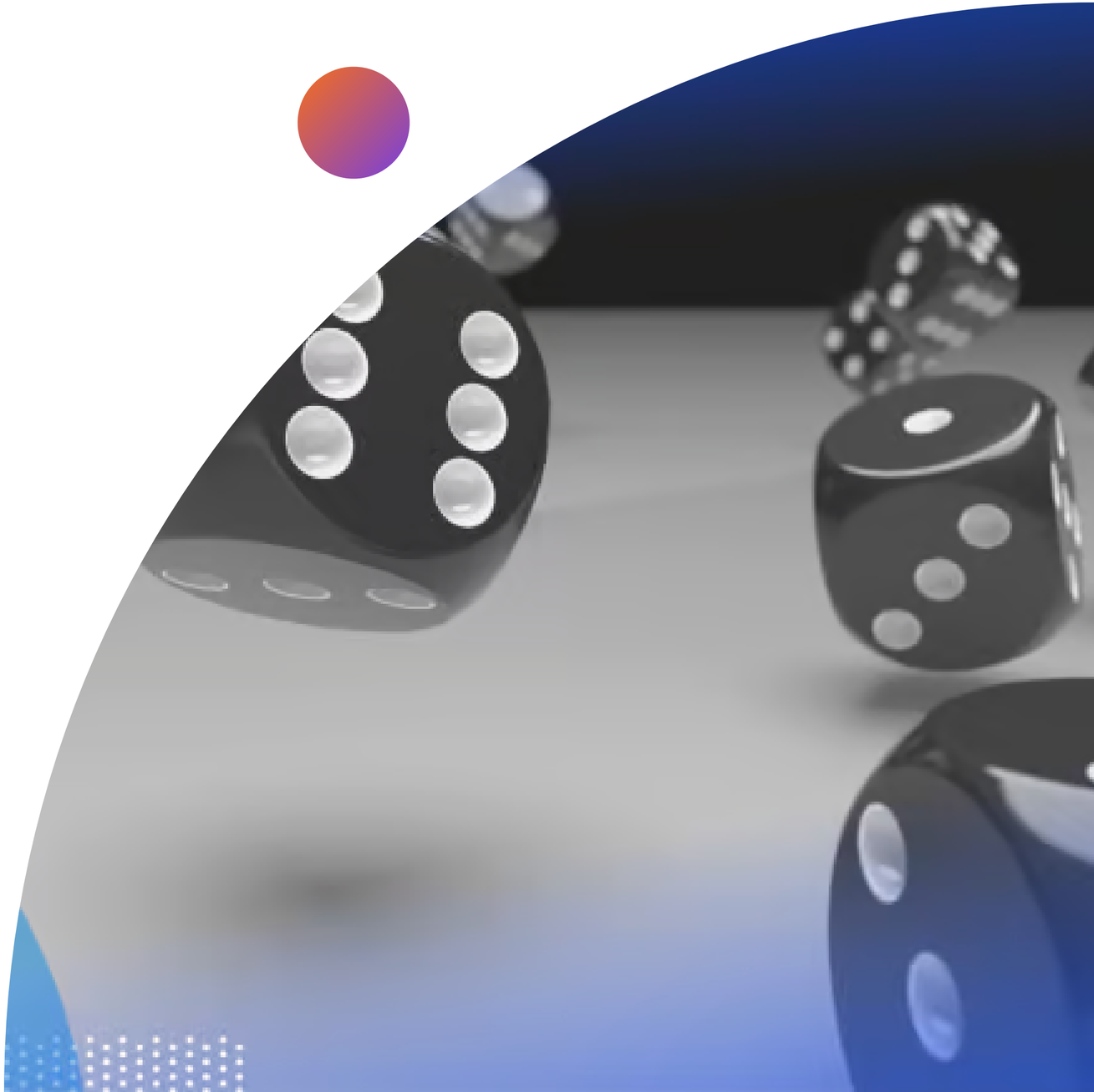
First Order Probability Theory

- Probabilistic Reasoning across the Causal Hierarchy:

Ibeling, Duligur, and Thomas Icard. "Probabilistic reasoning across the causal hierarchy." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 06, pp. 10170-10177. 2020.

- A Logic for Reasoning about Probabilities:

Fagin, Ronald, Joseph Y. Halpern, and Nimrod Megiddo. "A logic for reasoning about probabilities." Information and computation 87, no. 1-2 (1990): 78-128.



Probabilistic Reasoning across the Causal Hierarchy

a framework for probabilistic reasoning that takes into account the causal structure of a problem. The authors argue that traditional statistical methods, which focus on correlation rather than causation, are insufficient for many real-world problems where causal relationships play a crucial role.

Paper discusses the use of probability theory within the framework of causal graphs, which are used to represent the causal relationships between variables in a problem.

One of the key concepts discussed in the paper is the notion of probabilistic causality, which is captured by the axioms of probability theory. These axioms include:

1. **Non-negativity:** The probability of any event is non-negative.
2. **Normalization:** The probability of the entire sample space is equal to one.
3. **Additivity:** The probability of the union of two disjoint events is equal to the sum of their individual probabilities.
4. **Conditional probability:** The probability of an event given another event can be calculated using Bayes' rule.

A Logic for Reasoning about Probabilities

presents a logic for reasoning about probabilities, which extends first-order logic with probabilistic operators. The paper discusses several axioms that are used in this logic:

1. **Axiom of probability:** For any formula φ, the probability of φ is a real number between 0 and 1, inclusive.
2. **Axiom of tautology:** The probability of any tautology is 1.
3. **Axiom of negation:** The probability of the negation of a formula is 1 minus the probability of the formula.
4. **Axiom of conjunction:** The probability of a conjunction of two formulas is the product of their individual probabilities.
5. **Axiom of disjunction:** The probability of a disjunction of two formulas is the sum of their individual probabilities minus the probability of their conjunction.
6. **Axiom of conditional probability:** The probability of a formula φ given a formula ψ is the probability of the conjunction of φ and ψ divided by the probability of ψ.

These axioms are used to define the semantics of the probabilistic operators in the logic, including the conditional probability operator and the Bayesian update operator. The paper also discusses the soundness and completeness of the logic, as well as its limitations and possible extensions.

Definition 2.1. *A* **basic probability space** *is a triple* $(\overline{\Omega}, \overline{\mathbb{A}}, \overline{P})$ *where* $\overline{\Omega}$ *is a set,* $\overline{\mathbb{A}}$ *is a family of subsets of* $\overline{\Omega}$ *containing* $\overline{\Omega}$ *and* $\emptyset$*, and* $\overline{P}$ *is a real valued function on* $\overline{\mathbb{A}}$ *such that*

(a) $\overline{P}(\emptyset) = 0$*;*

(b) $\overline{P}(\overline{\Omega}) = 1$*;*

(c) for every $\overline{A} \subseteq \overline{B} \subseteq \overline{\Omega}$*,* $\overline{P}(\overline{A}) \leq \overline{P}(\overline{B})$*.*

Elements $\overline{A}$ *of* $\overline{\mathbb{A}}$ *are called basic events, and* $\overline{P}$ *is called basic probability.*

That these assumptions are not contradictory can be seen with $\overline{P}(\overline{A}) = \delta_{\overline{A},\overline{\Omega}}$ on any set $\overline{\Omega}$.

Definition 2.4. *A* **probability environment** *is a probability pre-environment* $((\Omega, \mathbb{A}, \mathcal{P}), \mathcal{R})$ *such that there exists a basic probability space* $(\overline{\Omega}, \overline{\mathcal{A}}, \overline{P})$ *satisfying all the requirements.*

What is Isabelle?

- a generic proof assistant and theorem prover
- provides a programming language called Isabelle/ML for defining formal theories, and a user interface for interactively constructing proofs using those theories.
- it has been used to formally verify the correctness of software and hardware designs, and to verify mathematical proofs. Isabelle is also extensible, allowing users to define their own formal theories and proof tactics.

Points of Reference

01

A formalization of metric spaces in HOL Light

Maggesi, Marco. "A formalization of metric spaces in HOL Light." Journal of Automated Reasoning 60, no. 2 (2018): 237-254.

02

Probability Theory in Isabelle/HOL

Hölzl, Johannes. "6.5. Probability Theory in Isabelle/HOL." In Joint Workshop of the German Research Training Groups in Computer Science, p. 100. GITO mbH Verlag, 2011.

01

A formalization of metric spaces in HOL Light

Maggesi, Marco. "A formalization of metric spaces in HOL Light." Journal of Automated Reasoning 60, no. 2 (2018): 237-254.

1. $d(x, y) \geq 0$,
2. $d(x, y) = 0$ if and only if $x = y$,
3. $d(x, y) = d(y, x)$,
4. $d(x, z) \leq d(x, y) + d(y, z)$.

```
let is_metric = new_definition
  `is_metric(d:X×X->real) ⟺
   (∀x y. &0 <= d(x,y)) ∧
   (∀x y. d(x,y) = &0 ⟺ x = y) ∧
   (∀x y. d(x,y) = d(y,x)) ∧
   (∀x y z. d(x,z) <= d(x,y) + d(y,z))`;;
```

02

Probability Theory in Isabelle/HOL

Hölzl, Johannes. "6.5. Probability Theory in Isabelle/HOL." In Joint Workshop of the German Research Training Groups in Computer Science, p. 100. GITO mbH Verlag, 2011.

The formalization in Isabelle includes definitions and theorems related to

- measurable spaces,
- measure functions,
- probability spaces,
- random variables,
- conditional probability,
- expectation,

and other topics in probability theory.

Probability Exercise

Example 4. *On a given day, casting of concrete structural elements at a construction project depends on the availability of material. The required material may be produced at the job site or delivered from a premixed concrete supplier. However, it is not always certain that these sources of material will be available. Furthermore, whenever it rains at the site, casting cannot be performed. On a given day, define the following elements:*

$E_1 =$ *there will be no rain*

$E_2 =$ *production of concrete material at the job site is feasible*

$E_3 =$ *supply of premixed concrete is available*

with the following respective probabilities: $P(E_1) = 0,8$, $P(E_2) = 0,7$, $P(E_3) = 0,95$ *and* $P(E_3|E_2^c) = 0,6$ *whereas* E_2 *and* E_3 *are statistically independent of* E_1.

(a) *Identify the following events in terms of* E_1, E_2, *and* E_3:

(i) $A =$ *casting of concrete elements can be performed on a given day;*

(ii) $B =$ *casting of concrete elements cannot be performed on a given day.*

(b) *determine the probability of the event* B.

(c) *If production of concrete material at the job site is not feasible, what is the probability that casting of concrete elements can still be performed on a given day?*

Probability Exercise. Solution

(a) $A = E_1 \cap (E_2 \cup E_3)$ $B = A^c = E_1^c \cup (E_2^c \cap E_3^c)$;

(b) by the independence of (any combination of) E_2, E_3 from E_1 we have

$$\begin{aligned} P(B) &= 1 - P(A) = 1 - P(E_1 \cap (E_2 \cup E_3)) \\ &= 1 - P(E_1)P(E_2 \cup E_3). \end{aligned}$$

Since $E_2 \cup E_3 = E_2 \cup (E_2^c \cap E_3)$ and $P(E_2^c \cap E_3) = P(E_3|E_2^c)P(E_2^c) = 0,6 \times (1 - 0,7) = 0,18$, we have

$$P(B) = 1 - (0,8 \times (0,7 + 0,18)) = 1 - 0,704 = 0.296. \quad (16)$$

(c) we have

$$\begin{aligned} P(A|E_2^c) &= \frac{P(A \cap E_2^c)}{P(E_2^c)} \\ &= \frac{P(E_1 \cap (E_2 \cup E_3) \cap E_2^c)}{P(E_2^c)} \\ &= \frac{P(E_1 \cap E_3 \cap E_2^c)}{P(E_2^c)} \\ &= \frac{P(E_1)P(E_3 \cap E_2^c)}{P(E_2^c)} = \frac{0,8 \times 0,18}{0.3} = 0.48 \end{aligned}$$

Contradiction

$$P(E_3 \cap E_2^c) = P(E_3 | E_2^c) P(E_2^c) = 0.18$$

$$P(E_3 \cap E_2) = P(E_3) - P(E_3 \cap E_2^c) = 0.77 > 0.7 = P(E_2)$$

Methodology:

01

Define the basic concepts and axioms of probability theory in Isabelle's logical language. For example, one can define the probability of an event as a real-valued function that satisfies the axioms of probability theory, such as additivity and non-negativity.

02

Define the basic operations and properties of random variables in Isabelle. For instance, one can define the expected value and variance of a random variable and prove their basic properties using Isabelle's proof tools.

03

Develop the theory of probability distributions in Isabelle. This involves defining various types of probability distributions, such as the normal distribution or the Poisson distribution, and proving their basic properties, such as their mean, variance, and moment-generating functions.

04

Use Isabelle's proof tools to prove important theorems in probability theory, such as the law of large numbers, the central limit theorem, and Bayes' theorem, and to check consistency of given assumptions.

Thank You