Extending and Automating Basic Probability Theory with Propositional Computability Logic

Keehang Kwon
khkwon@dau.ac.kr

Abstract: Classical probability theory is formulated using sets. In this paper, we extend classical probability theory with propositional computability logic[1] (CoL). Unlike other formalisms, computability logic is built on the notion of events/games, which is central to probability theory. The probability theory based on CoL is therefore useful for automating uncertainty reasoning. We describe some basic properties of this new probability theory.

1 Introduction

Classical probability theory[2] is formulated using sets. Unfortunately, the language of sets lacks expressiveness and is, in a sense, a low-level ‘assembly language’ of the probability theory. In this paper, we develop a ‘high-level approach’ to classical probability theory with propositional computability logic[1] (CoL). Unlike other formalisms such as sets, logic and linear logic, computability logic is built on the notion of events/games, which is central to probability theory. Therefore, CoL is a perfect place to begin the study of automating probability theory.

To be specific, CoL is well-suited to describing complex (sequential/parallel) experiments and events, and more expressive than set operations. In contrast, classical probability theory – based on $\cap$, $\cup$, etc – is designed to represent mainly the simple/additive events – the events that occur under a single experiment.

Naturally, we need to talk about composite/multiplicative events – events that occur under two different experiments. Developing probability along this line requires a new, powerful language. For example, consider the following events $E_1, E_2$:

$E_1$: toss a coin two times (events 1 and 2) and get H,T in that order.
$E_2$: toss two dices (which we call 1, 2) and get at least one 5.
Suppose a formalism has the notion of $\triangle$, $\triangledown$ (sequential-and/or) and $\land$, $\lor$ (parallel-and/or). Then $E_1$ would be written as $H^d \triangle T^d$. Similarly, $E_2$ would be written concisely as $(5^1 \lor 5^2)$. The formalism of classical probability theory fails to represent the above events in a concise way.

Computability logic\(^\text{[1]}\) provides a formal and consistent way to represent a wide class of experiments and events. In particular, multiplicative experiments (sequential and parallel experiments) as well as additive experiments (choice-AND, choice-OR) can be represented in this formalism.

\section{The Event Formulas}

The language is a variant of the propositional computability logic. For simplicity, we will not include sequential operators. The class of \textit{event formulas} is described by $E$-formulas given by the syntax rules below:

\[
E ::= a(t_1, \ldots, t_n) | \neg E | E \sqcup E | E \sqcap E | (E | E) \\
| E \parallel E | E \land E | E \lor E
\]

In the above, $a(t_1, \ldots, t_n)$ represents an atomic event called proposition where each $t_i$ is a term. For example, $d(6)$ represent the event where we get 6 from tossing a dice $d$. For readability, we often write $6^d$ (or simply 6) instead of $d(6)$.

Often events can be composed through the use of logical connectives. We will describe the definitions of these connectives.

First, we assume here experiments to be generated in stages. Thus our \textit{sample space} is a tree diagram of the form

\[
\{(a_1, \ldots, a_n), (b_1, \ldots, b_n), \ldots\}
\]

where each $a, b$ is an atomic event. As we shall see, this can also be represented as an event formula in \textit{set normal form} of the form

\[
(a_1 \land \ldots \land a_n) \sqcup (b_1 \land \ldots \land b_n) \sqcup \ldots
\]

In the sequel, we use the above two forms interchangeably.

An \textit{event space} is a subset of the sample space. In the sequel, we introduce a mapping $\ast$ which converts an event formula $E$ to an event space
\{E_1, \ldots, E_n\} of mutually exclusive points. This mapping makes it much
easier to compute the probability of \(E\). That is, \(p(E) = p(E_1) + \ldots + p(E_n)\).

To mentioned earlier, \(a(t)\) represents an atomic event. In addition,
\[
a(t)^* = \{a(t)\}
\]

The event \(-E\) represents the complement of the event \(E\) relative to the
universe \(U\).
\[
(-A)^* = U - A^*
\]

The choice-OR event \(A \sqcup B\) represents the event in which only one of the
event \(A\) and event \(B\) happen under a single experiment. For example, \(4^d \sqcup 5^d\)
represents the event that we get either 4 or 5 when a dice \(d\) is tossed. This
operation corresponds to the set union operation.
\[
(A \sqcup B)^* = A^* \cup B^*
\]

The choice-AND event \(A \sqcap B\) represents the event in which both event \(A\)
and event \(B\) happen under a single experiment. For example, \((2 \sqcup 4 \sqcup 6) \cap (2 \sqcup 3 \sqcup 5)\)
represents the event that we get both an even number and a prime number
in a single coin toss. This operation corresponds to the set intersection op-
eration.
\[
(A \sqcap B)^* = A^* \cap B^*
\]

The additive conditional event \(A|B\) represents the dependency between
\(A\) and \(B\): the event in which the event \(A\) happens given \(B\) has occurred
under a single experiment.

We can generalize the definition of events to joint events to deal with
multiple experiments. Here are some definitions.
The parallel-AND event \(A \land B\) represents the event in which both event \(A\)
and event \(B\) occur under two different experiments. For example, \((H^1 \land T^2) \sqcup (H^2 \land T^1)\)
represents the event that we get one head and one tail when two coins are
tossed. It is defined by the following:
\[
(A \land B)^* = A^* \times B^*
\]

Here, the (flattened) Cartesian conjunction of two sets \(A\) and \(B\), \(A \times B\)
is the following:
\[ A \times B = \{(a_1, \ldots, a_m, b_1, \ldots, b_n) | (a_1, \ldots, a_m) \in A \text{ and } (b_1, \ldots, b_n) \in B\} \]

For example, let \( A = \{(0, 1), (1, 2)\} \) and \( B = \{0, 1\} \). Then \( A \times B = \{(0, 1, 0), (0, 1, 1), (1, 2, 0), (1, 2, 1)\} \).

The parallel-OR event \( A \lor B \) represents the event in which at least one of event \( A \) and event \( B \) happen under two different experiments. For example, \(((4\sqcup 5) \lor (4\sqcup 5^2))\) represents the event that we get at least one 4 or one 5 when two dices are tossed. Formally,

\[ (A \lor B) =_{def} (A \land B) \sqcup (\neg A \land B) \sqcup (A \land \neg B) \]

The parallel conditional event \( A \parallel B \) (usually written as \( B \rightarrow A \) in logic) represents the event in which the event \( A \) happens given \( B \) has occurred before/after \( A \) under two experiments.

The following theorem substantially extends the traditional probability theory with new properties. These new properties are obtained by considering relative frequencies.

**Theorem 2.1** Let \( A \) be an event. Then the following properties hold.

1. \( p(\neg A) = 1 - p(A) \) % complement of \( A \)

2. \( p(A \lor B) = p(A) + p(B) - p(A \land B) \) % choice-or
   
   For example, \( H \lor T \) represents the event that \( H \) or \( T \) occur in a single coin toss. Now, it is easy to see that \( p(H \lor T) = 1 \).

3. \( p(A \land B) = p(A)p(B|A) = p(B)p(A|B) \) % choice-and
   
   For example, \( H \land T \) represents the event that \( H \) and \( T \) occur in a single coin toss. Now, \( p(H \land T) = 0 \). Note also that \( p(H \land H) = p(H) \). \( p(A|B) \) is the conditional probability of \( B \) given \( A \) in a single experiment. For example, \( H|T \) represents the event that \( H \) occurs given \( T \) occurs in a single coin toss. Now, \( p(H|T) = 0 \).

4. \( p(A \lor B) = p(A \land B) + p(\neg A \land B) + p(A \land \neg B) = 1 - p(\neg A \land \neg B) \) % parallel-or
   
   For example, \( p(H \lor 6) = 1 - p(T \land (1 \lor 2 \lor 3 \lor 4 \lor 5)) = 1 - 10/24 = 14/24 \).
(5) \( p(A \land B) = p(A)p(B \parallel A) = p(B)p(A \parallel B) \) % parallel-and

For example, suppose two coins are tossed. Now, \( p(H^1 \land H^2) = 1/4 \).
In addition, suppose a coin and a dice are tossed simultaneously. Then
\( p(H \land 6) = p(H)p(6) = 1/12 \).

% Below, \( p \) computes the probability of an event space rather than an event formula.

(6) \( p(A) = p(A^*) \) % \( A^* \) is the event space of \( A \).

For example, \( ((3 \sqcup 4) \cap 4)^* \) is \{4\}.

(7) \( p(\{E_1, \ldots, E_n\}) = p(\{E_1\}) + \ldots + p(\{E_n\}) \) % event space with \( n \) mutually exclusive elements.

(8) \( p(\{(a_1, \ldots, a_n)\}) = p(a_1 \land \ldots \land a_n) \) % event space with single element.
Here, each \( a_i \) is an atomic event.

In the above, we list some properties of our probability theory. In addition, we sometimes need to deal with sharing experiments among two experiments. Sharing experiments represents experiments which have been completed. Mathematical facts such as \( prime(2), odd(3) \) are such examples.

For this reason, we need to include the following:

\[ p(A \land A) = p(A \lor A) = p(A). \]

For example, let \( A \) be a (uncertain) statement that there are aliens in Vega. Then it is easy to see that these rules hold, as every occurrence of \( A \) represents the same event here.

3 Examples

Let us consider the following event \( E \) where \( E = \) roll a dice and get 4 or 5. The probabilities of \( E \) and \( E \lor E \) is the following:

\[ p(E) = p(4 \lor 5) = p(4) + p(5) = 1/3 \]

\[ p(E_1 \lor E_2) = 1 - (2/3 \times 2/3) = 5/9. \]
As another example, \((H^1 \land H^2) \sqcup (H^1 \land T^2) \sqcup (T^1 \land H^2)\) represents the event that at least one head comes up when two coins are tossed. Now, it is easy to see that \(p((H^1 \land H^2) \sqcup (H^1 \land T^2) \sqcup (T^1 \land H^2)) = \frac{3}{4}\).

As the last example, suppose two dice are tossed and we get 6 from one dice. What is the probability that we get 5 from another dice? This kind of problem is very cumbersome to represent/solve in classical probability theory. Fortunately, it can be represented/solved from the above formula in a concise way. It is shown below:

% computing the following probability requires converting the event to its event space.

\[
p((6^1 \land 5^2) \sqcup (6^2 \land 5^1)) \cap (6^1 \lor 6^2)) = p((5, 6), (6, 5)) \cap \{(6, 1), \ldots, (6, 6), (1, 6), \ldots, (5, 6)\}
\]

\[
= p((6^1 \land 5^2) \cup (6^2 \land 5^1)) = 5 \land 6 + p(6 \land 5) = \frac{2}{36}
\]

% computing the following does not require converting the event to its event space.

\[
p(6^1 \lor 6^2) = p(6^1 \land 6^2) + p(6^1 \land 6^2) + p(6^1 \land 6^2) = \frac{11}{36}
\]

From these two, we obtain the following in a purely algorithmic way:

\[
p((6^1 \land 5^2) \sqcup (6^2 \land 5^1)) | (6^1 \lor 6^2)) = \frac{2}{11}
\]

4 Two Versions of the Bayes Rule

In this section, we raise questions related to the interpretation of \(\cap\) in the Bayes rule. Most textbooks interpret \(\cap\) in an ambiguous, confusing way: sometimes as \(\sqcap\), and as \(\land\) in others. This is problematic, especially in automating probabilistic inference.

In considering automation of probability, it is very problematic/cumberson to use the Bayes rule in its current form. Suppose \(A\) can be partitioned into
k disjoint events $A_1, \ldots, A_k$. Understanding $\cap$ as $\sqcap$, the Bayes rule can be written as:

$$p(A_i | B) = \frac{p(A_i \cap B)}{p(A_1 \cap B) + \ldots + p(A_k \cap B)}$$

or

$$p(A_i | B) = \frac{p(A_i \cap B)}{p(A_1) p(B | A_1) + \ldots + p(A_k) p(B | A_k)}$$

This rule can easily be generalized to $\land$ as follows:

$$p(A_i \parallel B) = \frac{p(A_i \land B)}{p(A_1 \land B) + \ldots + p(A_k \land B)}$$

or

$$p(A_i \parallel B) = \frac{p(A_i \land B)}{p(A_1) p(B \parallel A_1) + \ldots + p(A_k) p(B \parallel A_k)}$$

That is, we need two versions of the Bayes rule and it is crucial to apply the correct version to get the correct answer. As a well-known example of the Bayes rule, consider the problem of sending 0 or 1 over a noisy channel. Let $r(0)$ be the event that a 0 is received. Let $t(0)$ be the event that a 0 is transmitted. Let $r(1)$ be the event that a 1 is received. Let $t(1)$ be the event that a 1 is transmitted. Now the question is: what is the probability of 0 having been transmitted, given 0 is received? In solving this kind of problem, it is more natural to use Bayes rule on $\land$, rather than on $\cap$. This is so because $p(t(0) \cap r(0)) = p(t(0) \cap r(0)) = 0$.

5 Conclusion

Computability logic [1] provides a formal and consistent way to represent a wide class of experiments and events. For this reason, we believe that probability theory based on computability logic is an interesting alternative to the traditional probability theory and uncertainty reasoning.
References

[1] G. Japaridze, “Propositional computability logic II”, ACM Transaction on Computational Logic, vol.7(2), pp.331–362, 2006.

[2] R.D. Yates and D.J.Goodman, “Probability and Stochastic Processes”, Wiley, 1999.