Propositional Probability Theory

Liron Karpati

Submission for Independent Study at Hebrew University of Jerusalem.
Advised by Dr. Asaf Weinstein.

Abstract

In expositions of Probability Theory, probability functions are defined either on sets of a sigma algebra or on propositions of a propositional language. This paper advocates in favor of defining probability functions on propositions. To this end, it will be shown that probability functions on propositions are natural (invariant preserving) generalizations of evaluation functions from propositional logic, giving an elegant grounding and foundation for the subject. In addition, we show that elementary outcomes can be understood as special kinds of propositions, therefore making a propositional approach to probability theory more fundamental.

1 Introduction

There are generally two main frameworks for setting up probability theory: probability on propositions (the propositional framework) and probability on sets of elementary outcomes (the set-theoretic framework). The set-theoretic framework is considered standard and is used by most textbooks, courses, and papers. Propositional frameworks are currently most often employed by Bayesian approaches to probability theory and epistemology. [1], [7], [5] Consequently, many statisticians, students, and practitioners are only familiar with the set theoretic framework for probability theory. As expressed in a classic textbook on the subject, "Kolmogorov's vision of founding probability theory on the concept of a normalized measure space has become the accepted orthodoxy." [2] The aim of this paper is to support a wider adoption of a propositional framework for probability theory by arguing that (i) a propositional framework founds probability theory as a natural extension of propositional logic and (ii) the elementary outcomes in the set-theoretic framework can be understood more fundamentally in terms of propositions. The paper is organized as follows. Section 2 outlines the set-theoretic framework as it is described in standard literature. Section 3.1 provides an introduction to propositional logic. Section 3.2 shows how we can naturally extend propositional logic to probability theory (and to a probabilistic logic). Section 4 compares the two frameworks and demonstrates the conceptual priority of a propositional framework over a set-theoretic framework.

In this paper, we will only consider finitely additive probability functions on finite probability spaces. Some might object that restriction to the finite case renders any theory developed of little value, given how many applications of probability theory require infinite spaces and countable additive measures. This is a fair objection. I would hardly make the claim that the propositional framework I will support in this paper is complete. It is far from it without a proper response to how to handle cases and applications that are currently handled by infinite spaces. Nevertheless restricting attention to the finite case can have considerable value. Is the theory of matrices of little value because it depends on restricting attention to finite vector spaces? Clearly not. In fact matrices helps one develop their intuition for linear operators in the infinite case. What is presented here is merely my initial steps at thinking through the foundations of probability theory.

2 Set-theoretic Framework

By a set-theoretic framework, I mean the approach to probability theory (usually done via measure theory) which begins by defining probability on sets of 'elementary outcomes' called 'events'. As expressed in a standard probability text by Bauer, in this framework, "the goal of probability theory is to provide methods of describing and analyzing experiments with random outcomes. In particular, mathematical models for an adequate study of such experiments involving chance have to be developed. In such experiments we are interested in the observation of 'events' or 'random magnitudes', as well as the calculation of the 'probability' with which such events occur." [2]

What are these 'events' and 'experiments'? In the set-theoretic framework, these notions are defined axiomatically via Kolomogorov's axioms. Bauer is explicit about the use of axiomatic foundations for the subject. "For the construction of a theory of probability intrinsic definitions of concepts like 'event' and 'probability' are not necessary, and in fact, to avoid logical difficulties and to give the theory the broadest and easiest applicability, such definitions are not worth attempting. Just as in the other areas of mathematics mentioned, in probability theory everything comes down to the formal properties of the concepts."[2] Essentially, the axioms used to give meaning to 'event' and 'experiment' are taken for granted because they give formal properties which seem to work. We will later see that the defining characteristics of probability functions (in a propositional framework) are supported by more than an 'it seems to work' argument and that we can in fact meaningful understand elementary outcomes (and consequently 'events' and 'experiments' which are made up of those elementary outcomes) in terms of propositions.

The standard axiomatic development is as follows. "Let Ω be a nonempty point set representing all possible outcomes of an experiment, and let Σ be an algebra of subset of Ω. The members of Σ, called events, are the collection of outcomes that are of interest to the experimenter." [8] By an algebra one means a set of subsets of Ω which is "nonempty and is closed under finite unions and complements. Let $P : \Sigma \to \mathbb{R}^+$ be a mapping, called a probability [function], defined for all elements of Σ so that the following rules are satisfied.

$$\text{For each } A \in \Sigma,\ 0 \le P(A) \text{ and } P(\Omega) = 1. \tag{1}$$

$$\text{If } A, B \in \Sigma,\ A \cap B = \emptyset \text{ then } P(A \cup B) = P(A) + P(B). \tag{2}$$

Such a P is called a 'finitely additive probability [function]." Sometimes the second axiom is given by a stronger "countable additivity" condition where if "$A_1, A_2, ...$ are disjoint events of Ω, such that $A = \bigcup_{k=1}^{\infty} A_k$ is also an event of Ω, then $P(A) = \sum_{k=1}^{\infty} P(A_k)$." [8] Because we are restricting our attention in this paper to finite sample spaces we only need finite additivity.

An alternative characterization is given in [2] via measure theory. "The basic context for defining probability-theoretic concepts is ... [a] normalized measure space... designated $(\Omega, \mathcal{A}, P)$, where Ω is a set, $\mathcal{A}$ a σ-algebra in Ω [previously denoted Σ] and P a measure on $\mathcal{A}$ normalized by $P(\Omega) = 1$."

3 Proposition-theoretic Framework

One of the important differences between the set-theoretic and propositional frameworks is that the axioms which characterize a probability function in the propositional framework are not justified on the ground of simply "giving the correct formal properties". Instead the axioms can be understood as generalizing classical propositional logic. The motivation for seeking such a generalization comes from the observation that in day-to-day life we do not really know what is in fact true/false. Instead, we often want to express a degree of certainty in statements and be able to correctly reason under a lack of complete certainty. The semantics of propositional logic only assigns two values True (1) and False (0) to propositions so it is natural to wonder if we can extend the semantics of propositional logic to instead assign graded values of belief to propositions.

3.1 Propositional Logic

3.1.1 Introductory Concepts

How do we go about extending the semantics of propositional logic to graded values? Let's first review the semantics of propositional logic. I use set notation here because it is the standard formalism, but one could use something closer to a type theory instead to avoid the infinite sets.

Definition 1. *A* ***finite propositional language*** $\mathcal{L}$ *consists of a set of atomic propositions*

$$Pr = \{A_1, ..., A_n\},$$

called ***atomic propositions****, some* ***connectives*** *(which we will restrict to)* $\neg, \wedge, \vee$*, and a set of* ***propositions****,* $\mathcal{F}$*, which is recursively defined in terms of atomic* *propositions and connectives like so:*

$$\begin{cases} A_1, ..., A_n \in \mathcal{F}, & \\ \neg\phi \in \mathcal{F} & \textit{if } \phi \in \mathcal{F}, \\ (\phi \wedge \psi) \in \mathcal{F} & \textit{if } \phi \in \mathcal{F} \textit{ and } \psi \in \mathcal{F}, \\ (\phi \vee \psi) \in \mathcal{F} & \textit{if } \phi \in \mathcal{F} \textit{ or } \psi \in \mathcal{F}. \end{cases} \tag{3}$$

Example 1. *Consider a scenario where we have three balls indexed* $1, 2, 3$ *in a bag and we draw from the bag. We can express this scenario with the atomic propositions* A_1, A_2, A_3 *where* A_i *means ball i was drawn. In this case,* $\neg A_1$ *means ball 1 was not drawn* $(A_2 \wedge A_3)$ *means ball 2 and ball 3 were both drawn,* $((A_1 \vee A_2) \vee A_3)$ *which we can just write as* $(A_1 \vee A_2 \vee A_3)$ *because of associativity of disjunctions means either ball 1 or ball 2 or ball 3 was drawn.*

In Example 1 we might want to be able to express that at least one of the balls must be drawn. In other words, that 'ball 1 or ball 2 or ball 3 was drawn' is true. In the semantics of propositional logic we use an evaluation function to map propositions to the values true (1) or false (0). An evaluation $v : \mathcal{F} \to \{0, 1\}$ is often characterized using a mix of natural language and mathematics like so:

$$v(\neg\phi) = \begin{cases} 1 & \text{if } v(\phi) = 0 \\ 0 & \text{if } v(\phi) = 1 \end{cases}$$
$$v(\phi \wedge \psi) = \begin{cases} 1 & \text{if } v(\phi) = 1 \text{ and } v(\psi) = 1 \\ 0 & \text{if } v(\phi) = 0 \text{ or } v(\psi) = 0 \end{cases}$$
$$v(\phi \vee \psi) = \begin{cases} 0 & \text{if } v(\phi) = 0 \text{ and } v(\psi) = 0 \\ 1 & \text{if } v(\phi) = 1 \text{ or } v(\psi) = 1 \end{cases}.$$

But, in using natural language, this is somewhat imprecise and as we will see, less illuminating. Instead we will characterize the same rules using mathematical operations.

Definition 2. *An* ***evaluation*** *is a function* $v : \mathcal{F} \to \{0, 1\}$ *which maps every proposition to its truth value and which satisfies the following conditions for propositions* ϕ, ψ*:*

$$v(\neg\phi) = 1 - v(\phi) \tag{4}$$
$$v(\phi \wedge \psi) = v(\phi) \cdot v(\psi) \tag{5}$$
$$v(\phi \vee \psi) = v(\phi) + v(\psi) - v(\phi) \cdot v(\psi) \tag{6}$$

A ***truth assignment*** *is the restriction of a valuation function to the set of atomic propositions,* $v|_{Pr} : \{A_1, ..., A_n\} \to \{0, 1\}$ *which maps atomic propositions to their truth value.*

We will denote the set of all evaluations for a finite propositional language as

$$V.$$

Once a truth assignment is specified, there is only one possible evaluation function which is consistent with it (given recursively by the conditions in Definition 2).

Proposition 1. *For any truth assignment* $v|_{Pr} : \{A_1, ..., A_n\} \to \{0, 1\}$*, there is exactly one evaluation* v *for which* $v(A_i) = v|_{Pr}(A_i)$ *for all* $A_i \in \{A_1, ..., A_n\}$*.*

Proof. We will not give a full proof here as it would require defining the complexity of a proposition which is not necessary for the purposes of the paper. The result intuitively follows Definition 2 because the only evaluation satisfying 4, 5, 6 would be recursively defined as followed:

$$\begin{cases} v(A_i) := v|_{Pr}(A_i) & \text{if } A_i \text{ is an atomic proposition,} \\ v(\neg\phi) := 1 - v(\phi) & \text{if } \phi \text{ is a proposition,} \\ v(\phi \wedge \psi) := v(\phi) \cdot v(\psi) & \text{if } \phi, \psi \text{ are propositions,} \\ v(\phi \vee \psi) := v(\phi) + v(\psi) - v(\phi) \cdot v(\psi) & \text{if } \phi, \psi \text{ are propositions.} \end{cases}$$

□

Usually evaluations are defined as extensions of truth assignments, but because probability functions will generalize evaluation functions, we choose to emphasize evaluations as their own object and a truth assignment as a restriction of an evaluation.

Example 2. *Continuing from Example 1, we can now make our intuition explicit. To express that* $(A_2 \wedge A_3)$ *can't ever happen (we can't draw ball 2 and ball 3 at the same time) we can write* $v(A_2 \wedge A_3) = 0$*. To express that one of the balls must be drawn we can write* $v(A_1 \vee A_2 \vee A_3) = 1$*.*

An important tool in the propositional logic toolkit is truth tables. Truth tables are a compact way to consider all possible truth-assignments (and consequently all possible evaluations). The first n columns give all the possible combinations of truth values for the atomic propositions and the later columns give the evaluation of particular propositions under those different combinations.

Example 3. *If we have three atomic propositions* A_1, A_2, A_3 *then the truth table would be:*

A_1	A_2	A_3	$A_2 \wedge A_3$	$A_1 \vee A_2 \vee A_3$	$\neg A_2 \vee \neg A_3$	$\neg(A_2 \wedge A_3)$
1	*1*	*1*	*1*	*1*	*0*	*0*
1	*1*	*0*	*0*	*1*	*1*	*1*
1	*0*	*1*	*0*	*1*	*1*	*1*
1	*0*	*0*	*0*	*1*	*1*	*1*
0	*1*	*1*	*1*	*1*	*0*	*0*
0	*1*	*0*	*0*	*1*	*1*	*1*
0	*0*	*1*	*0*	*1*	*1*	*1*
0	*0*	*0*	*0*	*0*	*1*	*1*

The 6th row corresponds to the evaluation, $v(A_1) = 0$, $v(A_2) = 1$, $v(A_3) = 0$, $v(A_2 \wedge A_3) = 0$ *and* $v(A_1 \vee A_2 \vee A_3) = 1$, $v(\neg A_2 \vee \neg A_3) = 0$, $v(\neg(A_2 \wedge A_3)) = 0$.

The purpose of propositional logic is not only to help us express scenarios like Example 1 formally, but rather we want to express scenarios formally so to be able to reason about them precisely. We want to be able to infer what we should believe from what we already assume. The concept we use to express valid inference in propositional logic is entailment.

Definition 3. *We say propositions (also called the* ***premises****)* $\phi_1, ..., \phi_k$ ***entail*** *proposition (also called the* ***conclusion****)* ψ*, denoted*

$$\phi_1, ..., \phi_k \models \psi,$$

exactly when for every evaluation v*,*

$$\textit{if } v(\phi_1) = 1, \ldots, v(\phi_k) = 1 \textit{ then } v(\psi) = 1. \tag{7}$$

We say two propositions ϕ *and* ψ *are* ***logically equivalent****, denoted*

$$\phi \equiv \psi,$$

exactly when they entail each other,

$$\phi \models \psi \textit{ and } \psi \models \phi. \tag{8}$$

Definition 4. *We call* τ *is a* ***tautology*** *exactly when it is true under every evaluation,*

$$\models \tau, \tag{9}$$

and we say $\perp$ *is a* ***contradiction*** *exactly when it is false under ever evaluation*

$$\models \neg\perp. \tag{10}$$

Example 4. *Truth tables help us identify entailments. Consider the Table from Example 3 and the following examples which we will interpret in the context of Example 1.*

1. *$A_1 \models A_1 \vee A_2 \vee A_3$ since if $v(A_1) = 1$ then by Definition 2, $v(A_1 \vee A_2 \vee A_3) = 1$. This makes sense because if we know we drew ball 1 then certainly we drew ball 1 or ball 2 or ball 3. The reverse, however, is not true!*
2. *$A_1 \vee A_2 \vee A_3 \not\models A_1$. We can see this from the truth table. There is a row in which $A_1 \vee A_2 \vee A_3$ is true but the A_1 is not. In particular, row 7 gives us the counter-example $v(A_1 \vee A_2 \vee A_3) = 1$ but $v(A_1) = 0$. Just because we draw one of the balls does not guarantee that we drew ball 1 since we could have drawn ball 3, for example.*
3. *$\neg A_2 \vee \neg A_3 \equiv \neg(A_2 \wedge A_3)$. We can see this from the truth table because the entries in both of their columns always match. In the context of the example, this means that to believe that we cannot draw ball 2 and ball 3 at the same time is the same as believing that either we did not draw ball three or we did not draw ball 2.*

3.1.2 Atomic States

The final concepts of propositional logic that we will need to introduce are atomic states and the disjunctive normal form of a proposition. We will first define them and then explain them in the context of the truth tables.

Definition 5. *In a propositional language with atomic propositions $A_1, ..., A_n$ an **atomic state** is a proposition of the form*

$$\pm A_1 \wedge \pm A_2 \wedge ... \wedge \pm A_n$$

where $+A_i$ stands in for A_i and $-A_i$ stands in for $\neg A_i$. We denote the set of atomic states by

$$\Omega.$$

Denote by

$$A^+(\omega), A^-(\omega)$$

the set of atomic propositions which appear without/with a negation (respectively) in the atomic state ω.

Example 5. *Let $\omega = A_1 \wedge \neg A_2 \wedge A_3 = +A_1 \wedge -A_2 \wedge +A_3$.*

$$A^+(\omega) = \{A_1, A_3\},$$
$$A^-(\omega) = \{A_2\}.$$

Proposition 2. *In a propositional language with atomic propositions $A_1, ..., A_n$, there are 2^n atomic states.*

Proof. This follows from a simple counting argument. For each $A_1, ..., A_n$ we either choose a $+$ or a $-$ sign. □

Definition 6. *In a propositional language with atomic propositions $A_1, ..., A_n$, we associate each atomic state ω with the evaluation*

$$v_\omega(A) := \begin{cases} 1 & A \in A^+(\omega) \\ 0 & A \in A^-(\omega) \end{cases}. \tag{11}$$

Technically (11) only defines a truth assignment but that truth assignment is uniquely extendable to an evaluation by Proposition 1.

Proposition 3. *In a propositional language with atomic propositions $A_1, ..., A_n$, the function $f : \Omega \to V$ mapping atomic states to evaluations by $\omega \mapsto v_\omega$ is a bijection.*

Proof. First we show that f is a function. Suppose $v_\omega = v_{\omega'}$. Then for every atomic state A_i, $v_\omega(A_i) = v_{\omega'}(A_i)$. By (11), this means $A^+(\omega) = A^+(\omega')$ and $A^-(\omega) = A^-(\omega')$. Since ω, ω' consist of the same conjunction of of atomic states (and their negations) they are the same atomic state, $\omega = \omega'$.
Now we show that f is a bijection. Suppose $\omega \neq \omega'$. Then, without loss of generality, there is some atomic state A such that $A \in A^+(\omega)$ and $A \notin A^+(\omega')$. Therefore $v_\omega(A) = 1$ and $v_{\omega'}(A) = 0$ so $v_\omega \neq v_{\omega'}$. This shows that f is one-to-one. By Definition 2, it is easy to see that there are 2^n possible truth assignments and so 2^n possible evaluations by Proposition 1. Using Proposition 2, $|\Omega| = |V|$, which means f must be a bijection. □

Proposition 4. $v_\omega(\omega') = 1$ *iff* $\omega' = \omega$.

Proof. Let

$$\omega = +A_1 \wedge ... \wedge +A_k \wedge -A_{k+1} \wedge ... \wedge -A_n,$$
$$\omega' = +A_{i_1} \wedge ... \wedge +A_{i_k} \wedge -A_{i_{k+1}} \wedge ... \wedge -A_{i_n}.$$

Clearly,

$$A^+(\omega) = \{A_1, \dots, A_k\} \text{ and } A^-(\omega) = \{A_{k+1}, \dots, A_n\}$$

and

$$A^+(\omega') = \{A_{i_1}, \dots, A_{i_k}\} \text{ and } A^-(\omega') = \{A_{i_{k+1}}, \dots, A_{i_n}\}.$$

Suppose $v_\omega(\omega') = 1$. Then by recursive application of Definition 2,

$$v_\omega(A_{i_1}) = 1, ..., v_\omega(A_{i_k}) = 1$$

and

$$v_\omega(\neg A_{i_{k+1}}) = 1, ..., v_\omega(\neg A_{i_n}) = 1$$

which means

$$v_\omega(A_{i_{k+1}}) = 0, ..., v_\omega(A_{i_n}) = 0.$$

By (11),

$$A^+(\omega) = \{A_{i_1}, \dots, A_{i_k}\} \text{ and } A^-(\omega) = \{A_{i_{k+1}}, \dots, A_{i_n}\}.$$

Therefore, $A^+(\omega) = A^+(\omega')$ and $A^-(\omega) = A^-(\omega')$, which means that $\omega = \omega'$.
Suppose $\omega = \omega'$. (11) and Definition 2 imply that $v_\omega(\omega) = 1$. □

Propositions 3 and 4 show us that there is a special bijection between evaluations and atomic states. This special bijection gives us a deeper understanding of the connection between evaluations and rows in the truth table. Each atomic state is true in exactly one row (evaluation). We can read of which atomic state is true in a row by looking at the first n columns. An example will make this clearer.

Example 6. *Consider the truth table:*

atomic state	A_1	A_2	A_3	$A_2 \wedge A_3$	$A_1 \vee A_2 \vee A_3$	$\neg A_2 \vee \neg A_3$	$\neg(A_2 \wedge A_3)$
$A_1 \wedge A_2 \wedge A_3$	1	1	1	1	1	0	0
$A_1 \wedge A_2 \wedge \neg A_3$	1	1	0	0	1	1	1
$A_1 \wedge \neg A_2 \wedge A_3$	1	0	1	0	1	1	1
$A_1 \wedge \neg A_2 \wedge \neg A_3$	1	0	0	0	1	1	1
$\neg A_1 \wedge A_2 \wedge A_3$	0	1	1	1	1	0	0
$\neg A_1 \wedge A_2 \wedge \neg A_3$	0	1	0	0	1	1	1
$\neg A_1 \wedge \neg A_2 \wedge A_3$	0	0	1	0	1	1	1
$\neg A_1 \wedge \neg A_2 \wedge \neg A_3$	0	0	0	0	0	1	1

The atomic state $A_1 \wedge A_2 \wedge A_3$ is true only in the first row since that is the only case in which all the atomic propositions are true. $A_1 \wedge A_2 \wedge \neq A_3$ is true only in the second row since that is the only case in which A_1, A_2 are true and A_3 false ($\neg A_3$ true). We can therefore label each row with the atomic state which it uniquely makes true.

Proposition 5. *In a propositional language with atomic propositions $A_1, ..., A_n$, for each atomic state ω, $v_\omega(\phi) = 1$ iff $\omega \models \phi$.*

Proof. ($\rightarrow$) Suppose $v_\omega(\phi) = 1$. To show that $\omega \models \phi$, we must prove that all evaluations under which ω is true, also make ϕ true. This is trivial because by Proposition 4 and Definition 6 we know there is only one evaluation which makes ω true, v_ω, and by assumption $v_\omega(\phi) = 1$. Therefore $\omega \models \phi$.
($\leftarrow$) Suppose $\omega \models \phi$. Then any for any evaluation v for which $v(\omega) = 1$ it must also be the case that $v(\phi) = 1$. v_ω is one such evaluation, so $v_\omega(\phi) = 1$. □

Theorem 6. *In a propositional language with atomic propositions $A_1, ..., A_n$, every proposition ϕ is logically equivalent to a unique disjunction of atomic states, called the* ***disjunctive normal form*** *for the proposition, given by*

$$\bigvee_{\omega \models \phi} \omega.$$

Proof. Let ϕ be a proposition.
Existence: We will show that ϕ and $\bigvee_{\omega \models \phi} \omega$ are logically equivalent.
($\rightarrow$) Suppose $v(\phi) = 1$. By Proposition 4, there is some atomic state, ω, for which $v(\omega) = 1$. By Proposition 5, $\omega \models \phi$. Therefore by Definition 2, $v(\bigvee_{\omega \models \phi} \omega) = 1$. Finally, by Definition 3, $\phi \models \bigvee_{\omega \models \phi} \omega$.
($\leftarrow$) Suppose $v(\bigvee_{\omega \models \phi} \omega) = 1$. This means there must be some ω^*, such that $\omega^* \models \phi$ for which $v(\omega^*) = 1$. Since $\omega^* \models \phi$, it follows that $v(\phi) = 1$. Thus, $\bigvee_{\omega \models \phi} \omega \models \phi$.

By showing entailment in both direction, by Definition 3, we have shown logical equivalence. We now show uniqueness.

Uniqueness: Suppose $O \subseteq \Omega$ and $\bigvee_{\omega \in O} \omega$ is logically equivalent to ϕ. If there is a $\omega \in O$ for which $w \not\models \phi$ then under the evaluation v_ω, $v_\omega(\bigvee_{\omega \in O} \omega) = 1$ and $v_\omega(\phi) = 0$ which is a contradiction. If there is some ω for which $\omega \models \phi$ and which is not in O, then $v_\omega(\phi) = 1$ and because v_ω uniquely makes ω true and all other atomic states false, $v_\omega(\bigvee_{\omega \in O} \omega) = 0$. Again a contradiction. We see that O must consist of exactly the atomic states ω which entail ϕ, as desired. □

Example 7. *The disjunctive normal form of a proposition can be read off from a truth table. It will help to present a couple of explicit examples from which the pattern is made clear.*

1. $A_2 \wedge A_3 \equiv (A_1 \wedge A_2 \wedge A_3) \vee (\neg A_1 \wedge A_2 \wedge A_3)$

atomic state	A_1	A_2	A_3	$A_2 \wedge A_3$	$A_1 \vee A_2 \vee A_3$	$\neg A_2 \vee \neg A_3$	$\neg(A_2 \wedge A_3)$
$A_1 \wedge A_2 \wedge A_3$	1	1	1	1	1	0	0
$A_1 \wedge A_2 \wedge \neg A_3$	1	1	0	0	1	1	1
$A_1 \wedge \neg A_2 \wedge A_3$	1	0	1	0	1	1	1
$A_1 \wedge \neg A_2 \wedge \neg A_3$	1	0	0	0	1	1	1
$\neg A_1 \wedge A_2 \wedge A_3$	0	1	1	1	1	0	0
$\neg A_1 \wedge A_2 \wedge \neg A_3$	0	1	0	0	1	1	1
$\neg A_1 \wedge \neg A_2 \wedge A_3$	0	0	1	0	1	1	1
$\neg A_1 \wedge \neg A_2 \wedge \neg A_3$	0	0	0	0	0	1	1

2. $A_2 \equiv (A_1 \wedge A_2 \wedge A_3) \vee (A_1 \wedge A_2 \wedge \neg A_3) \vee (\neg A_1 \wedge A_2 \wedge A_3) \vee (\neg A_1 \wedge A_2 \wedge \neg A_3)$

atomic state	A_1	A_2	A_3	$A_2 \wedge A_3$	$A_1 \vee A_2 \vee A_3$	$\neg A_2 \vee \neg A_3$	$\neg(A_2 \wedge A_3)$
$A_1 \wedge A_2 \wedge A_3$	1	1	1	1	1	0	0
$A_1 \wedge A_2 \wedge \neg A_3$	1	1	0	0	1	1	1
$A_1 \wedge \neg A_2 \wedge A_3$	1	0	1	0	1	1	1
$A_1 \wedge \neg A_2 \wedge \neg A_3$	1	0	0	0	1	1	1
$\neg A_1 \wedge A_2 \wedge A_3$	0	1	1	1	1	0	0
$\neg A_1 \wedge A_2 \wedge \neg A_3$	0	1	0	0	1	1	1
$\neg A_1 \wedge \neg A_2 \wedge A_3$	0	0	1	0	1	1	1
$\neg A_1 \wedge \neg A_2 \wedge \neg A_3$	0	0	0	0	0	1	1

3.2 Probability as Extended Logic

In this section we will present how probability theory on propositions follows as a natural extension of propositional logic. The way this is usually done is via a representation theorem like Cox's Theorem [1, Ch. 1, 2] which gives some seemingly innocuous assumptions about what a probability function should satisfy and shows that those assumptions imply the rules of conditional probability theory like the product rule $P(A \wedge B \mid C) = P(A \mid B \wedge C)P(B \mid C)$. The reason Cox's Theorem is sometimes understood by people as showing that probability theory naturally extends logic is that part of the 'innocuous assumptions' are that conditional probability assignments are consistent with propositional logic. A precise account of this is given in [6]. One discomfort I have with Jaynes' foundations is that probability assignments are always conditional and that the logical environment of propositions is not made explicit [1, Ch. 1, 2].

The account that will be presented here is different. I will be showing that probability assignments naturally generalize evaluations. Crucially, probability assignments will not need to be conditional (just as evaluations are not conditional). I do not use a representation theorem but rather appeal to the aesthetic mathematical sense which appears to guide much of mathematics. My hope is that the reader will find the connections between evaluations and probability assignments motivating enough to call the leap from one to the other 'natural'.

3.2.1 The Structure Preserving Axioms

In this section I will present what I call the structure preserving axioms. To the extent that I have read the literature, I have not come across probability functions defined in terms of the structure preserving axioms. I will first present the axioms and then explain how they naturally generalize the properties of evaluation functions to the interval $[0, 1]$.

Definition 7. *A* ***probability function*** *is a function $p : \mathcal{F} \to [0, 1]$ which satisfies the following properties for propositions $\phi, \psi \in \mathcal{F}$:*
A1 $p(\neg\phi) + p(\phi) = 1$*;*
A2 $p(\phi \vee \phi) + p(\phi \wedge \psi) = p(\phi) + p(\psi)$*;*
A3 $p(\tau) = v(\tau)$ *for a tautology τ and evaluation v.*

The first two axioms follow naturally from invariants we can easily discover im Definition 2. Recall that by (5).

$$v(\phi \wedge \psi) = v(\phi) \cdot v(\psi)$$

and (6)

$$v(\phi \vee \psi) = v(\phi) + v(\psi) - v(\phi) \cdot v(\psi)$$

which are begging to be combined into the familiar form

$$v(\phi \vee \psi) = v(\phi) + v(\psi) - v(\phi \wedge \psi).$$

We can, then, rewrite the conditions in Definition 2 into the more symmetric forms:

$$v(\neg\phi) + v(\phi) = 1 \tag{12}$$
$$v(\phi \wedge \psi) = v(\phi) \cdot v(\psi) \tag{13}$$
$$v(\phi \vee \psi) + v(\phi \wedge \psi) = v(\phi) + v(\psi) \tag{14}$$

In this form, the *linear* conditions (12) and (14) also reveal to us something profound about the structure of propositional logic semantics: there is a duality between $\neg$ and its absence shown by (12) and another duality between $\wedge$ and $\vee$ shown by (14). If we replace v with p, therefore preserving these dualities for a probability function, we get axioms ***A1*** and ***A2*** respectively.

A3 comes from preserving a different invariant of evaluations. Notice that ***A3*** does not depend on the choice of evaluation. This is because the value of a tautology is another structural invariant of the evaluation function. ***A3*** codifies this invariance and grounds the values of a probability function in the values of an evaluation function. Of course, we could have equivalently specified ***A3*** in terms of contradictions.

Proposition 7. *For a contradiction $\perp \in \mathcal{F}$, $p(\perp) = 0$.*

Proof. Let $\perp$ be a contradiction. $\neg\perp$ is a tautology so by ***A1*** and ***A3***, $p(\perp) = 1 - p(\neg\perp) = 1 - 1 = 0$. □

Why can't we also generalize condition (13) and include $p(\phi \wedge \psi) = p(\phi) \cdot p(\psi)$ in the axioms? If we were to accept $p(\phi \wedge \psi) = p(\phi) \cdot p(\psi)$ as an axiom which holds for all $\phi, \psi \in \mathcal{F}$ then we would be forced to conclude that for any $\varphi \in \mathcal{F}$,

$$0 = p(\perp) = p(\varphi \wedge \neg\varphi) = p(\varphi) \cdot p(\neg\varphi)$$

so one of $p(\varphi), p(\neg\varphi)$ must be 0 and by ***A1*** the other must be 1. This would mean that every proposition can only take on the values $0, 1$ and a probability function would simply be an evaluation! Since our goal was to generalize evaluations to the interval $[0, 1]$ we cannot include a probabilistic version of (13) in our axioms. Of course, when $p(\phi \wedge \psi) = p(\phi) \cdot p(\psi)$ happens to be true of *some* propositions $\phi, \psi \in \mathcal{F}$, they are in a well known relationship called *probabilistic independence*.

It turns out that just by preserving the invariants of evaluation and extending them to a function on the interval $[0, 1]$, we recover a finitely additive probability function. That probability functions generalize evaluations can be seen most clearly from the following result.

Proposition 8. *An evaluation v is a probability function.*

Proof. We will show that v satisfies ***A1***, ***A2***, and ***A3***. By 4, for any proposition ϕ, $v(\neg\phi) + v(\phi) = 1$ (***A1***). Combining 5 and 6 we get that $v(\phi \vee \psi) + v(\phi \wedge \psi) = v(\phi) + v(\psi)$ (***A2***). ***A3*** follows since v is an evaluation already! □

We will now prove some results to demonstrate that our axioms are sufficiently powerful to characterize a finitely additive probability function. First we will *derive* that logically equivalent propositions have the same probability and then we will prove finite additivity.

Proposition 9. *For any propositions $\phi, \psi \in \mathcal{F}$, if $\phi \models \psi$ then $p(\phi) \leq p(\psi)$.*

Proof. Suppose $\phi \models \psi$. First we show that $\models \neg\phi \vee \psi$. If v is an evaluation such that $v(\phi) = 1$ then by Definition 3 $v(\psi) = 1$. So, by Definition 2, $v(\neg\phi \vee \psi) = 1$. On the other hand, if $v(\phi) = 0$, then by Definition 2, $v(\neg\phi) = 1$. So, $v(\neg\phi \vee \psi) = 1$. Therefore $\models \neg\phi \vee \psi$, or in words, $\neg\phi \vee \psi$ is a tautology.
By ***A3***, ***A2***, and ***A1***

$$\begin{aligned} p(\neg\phi \vee \psi) &= 1 \\ p(\neg\phi) + p(\psi) - p(\neg\phi \wedge \psi) &= 1 \\ 1 - p(\phi) + p(\psi) - p(\neg\phi \wedge \psi) &= 1 \\ p(\psi) &= p(\phi) + p(\neg\phi \wedge \psi) \\ p(\phi) &\leq p(\psi). \end{aligned}$$

□

Corollary 10. *For any propositions $\phi, \psi \in \mathcal{F}$, if $\phi \equiv \psi$ then $p(\phi) = p(\psi)$.*

Proof. if $\phi \equiv \psi$ then $\phi \models \psi$ and $\psi \models \phi$. So by Proposition 9, $p(\phi) = p(\psi)$. □

Definition 8. *Propositions ϕ and ψ are* ***disjoint*** *exactly when*

$$\phi \wedge \psi \equiv \bot.$$

Proposition 11. *Let p be a probability function on a finite propositional language $\mathcal{L}$. If ϕ and ψ are disjoint then*

$$p(\phi \vee \psi) = p(\phi) + p(\psi).$$

More generally, for pairwise disjoint propositions $\phi_1, \ldots, \phi_k$,

$$p(\bigvee_{i=1}^{k} \phi_i) = \sum_{i=1}^{k} p(\phi_i). \tag{15}$$

Proof. We prove this by induction.
Base case: By ***A2***, $p(\phi \vee \psi) = p(\phi) + p(\psi) - p(\phi \wedge \psi)$. Since $\phi \wedge \psi \equiv \bot$, by Proposition 7, $p(\phi \wedge \psi) = 0$. Therefore $p(\phi \vee \psi) = p(\phi) + p(\psi)$.
Inductive Case: Suppose (15) holds for $k \in \mathbb{N}$ and that we have $\phi_1, \ldots, \phi_{k+1}$ pairwise disjoint propositions. Then by the base case and the inductive hypothesis:

$$p(\bigvee_{i=1}^{k+1} \phi_i) = p(\bigvee_{i=1}^{k} \phi_i \vee \phi_{k+1})$$

$$= p(\bigvee_{i=1}^{k} \phi_i) + p(\phi_{k+1})$$
$$= \sum_{i=1}^{k} p(\phi_i) + p(\phi_{k+1})$$
$$= \sum_{i=1}^{k+1} p(\phi_i)$$

□

To summarize, a probability function is a natural generalization of evaluations to the interval $[0, 1]$ in that it is defined by simply preserving the structural invariants of evaluations.

3.2.2 Insight into Probability Functions

Having defined probability functions on propositions, in this section we will examine an important result about probability on propositions which shows that all probability values are determined by the probability values of the atomic states. It is so important, that it is taken to be axiomatic in [5]. First we must prove that atomic states are disjoint.

Proposition 12. *Let $\mathcal{L}$ be a finite propositional language with atomic states Ω. Any two distinct atomic states $\omega, \omega' \in \Omega$ are disjoint.*

Proof. Suppose ω, ω' differ in atomic proposition A_i. Then $\omega \wedge \omega' \equiv A_i \wedge \neg A_i \wedge (\dots) \equiv \perp$. □

Theorem 13. *Let p be a probability function on a finite propositional language $\mathcal{L}$, with set of atomic states Ω. For any proposition $\phi \in \mathcal{F}$,*

$$p(\phi) = \sum_{\omega \models \phi} p(\omega)$$

where the sum ranges over those $\omega \in \Omega$ that entail ϕ.

Proof. By Theorem 6, $\phi \equiv \bigvee_{\omega \models \phi} \omega$. Because atomic states are pairwise disjoint by Proposition 12, we can apply Proposition 11,

$$p(\phi) = p(\bigvee_{\omega \models \phi} \omega) = \sum_{\omega \models \phi} p(\omega).$$

□

We can see from Theorem 13 that once the probability of the atomic states are specified, the probability of all other propositions are determined. This allows us to augment the truth table to be able to read off the probability of any proposition.

Example 8. *We add a column for the probability of each atomic state.*

atomic state	p	A_1	A_2	A_3	$A_2 \wedge A_3$	$A_1 \vee \neg A_1$	$\neg A_2 \vee \neg A_3$
$A_1 \wedge A_2 \wedge A_3$	0.1	*1*	*1*	*1*	*1*	*1*	*0*
$A_1 \wedge A_2 \wedge \neg A_3$	0	*1*	*1*	*0*	*0*	*1*	*1*
$A_1 \wedge \neg A_2 \wedge A_3$	0.2	*1*	*0*	*1*	*0*	*1*	*1*
$A_1 \wedge \neg A_2 \wedge \neg A_3$	0.3	*1*	*0*	*0*	*0*	*1*	*1*
$\neg A_1 \wedge A_2 \wedge A_3$	0	*0*	*1*	*1*	*1*	*1*	*0*
$\neg A_1 \wedge A_2 \wedge \neg A_3$	.15	*0*	*1*	*0*	*0*	*1*	*1*
$\neg A_1 \wedge \neg A_2 \wedge A_3$	.25	*0*	*0*	*1*	*0*	*1*	*1*
$\neg A_1 \wedge \neg A_2 \wedge \neg A_3$	0	*0*	*0*	*0*	*0*	*1*	*1*

We can read off the probability of propositions based on the probabilities of the atomic states. Recall Example 7 for the first two disjunctive normal forms.

- $p(A_2 \wedge A_3) = p(A_1 \wedge A_2 \wedge A_3) + p(\neg A_1 \wedge A_2 \wedge A_3)$
  $= 0.1 + 0$
  $= 0.1$
- $p(A_2) = p(A_1 \wedge A_2 \wedge A_3) + p(A_1 \wedge A_2 \wedge \neg A_3) + p(\neg A_1 \wedge A_2 \wedge A_3) + p(\neg A_1 \wedge A_2 \wedge \neg A_3)$
  $= 0.1 + 0 + 0 + 0.15$
  $= 0.25$
- $p(A_1 \vee \neg A_1) = p(A_1 \wedge A_2 \wedge A_3) + p(A_1 \wedge A_2 \wedge \neg A_3) + p(A_1 \wedge \neg A_2 \wedge A_3)$
  $+ p(A_1 \wedge \neg A_2 \wedge \neg A_3) + p(\neg A_1 \wedge A_2 \wedge A_3) + p(\neg A_1 \wedge A_2 \wedge \neg A_3)$
  $+ p(\neg A_1 \wedge \neg A_2 \wedge A_3) + p(\neg A_1 \wedge \neg A_2 \wedge \neg A_3)$
  $= 1$

Because the probability of any proposition is fully specified by the probability of the atomic states, we can rewrite Theorem 13 in terms of vectors.

Definition 9. *For a propositional language $\mathcal{L}$, with set of atomic states $\Omega = \{\omega_1, ..., \omega_{2^n}\}$, and equipped with a probability function p, denote the vector of probabilities of the atomic states by*

$$\vec{p} = [p(\omega_1), ..., p(\omega_{2^n})]^T$$

Further for each proposition $\phi \in \mathcal{F}$, let $\vec{I_\phi} = [I_{\phi 1}, ..., I_{\phi 2^n}]^T$ be the indicator vector such that

$$I_{\phi i} = \begin{cases} 1 & \text{if } \omega_i \models \phi \\ 0 & \text{otherwise} \end{cases}.$$

Corollary 14. *For a probability function p on a finite propositional language $\mathcal{L}$,*

$$p(\phi) = \vec{I_\phi} \cdot \vec{p}$$

for proposition $\phi \in \mathcal{F}$.

Proof. This follows from Theorem 13 and Definition 9. □

Example 9. *Continuing Example 8, label the atomic states in descending order of the truth table:*

$$\begin{aligned}
&\omega_1 = A_1 \wedge A_2 \wedge A_3, \;\; \omega_2 = A_1 \wedge A_2 \wedge \neg A_3, \;\; \omega_3 = A_1 \wedge \neg A_2 \wedge A_3 \\
&\omega_4 = A_1 \wedge \neg A_2 \wedge \neg A_3, \;\; \omega_5 = \neg A_1 \wedge A_2 \wedge A_3, \;\; \omega_6 = \neg A_1 \wedge A_2 \wedge \neg A_3 \\
&\omega_7 = \neg A_1 \wedge \neg A_2 \wedge A_3, \;\; \omega_8 = \neg A_1 \wedge \neg A_2 \wedge \neg A_3
\end{aligned}$$

In Example 8,

$$\begin{aligned}
&\vec{p} = [0.1,\ 0,\ 0.2,\ 0.3,\ 0,\ 0.15,\ 0.25,\ 0]^T \\
&\vec{I}_{A_2 \wedge A_3} = [1, 0, 0, 0, 1, 0, 0, 0]^T \\
&\vec{I}_{A_2} = [1, 1, 0, 0, 1, 1, 0, 0]^T \\
&\vec{I}_{\tau} = [1, 1, 1, 1, 1, 1, 1, 1]^T
\end{aligned}$$

It can be easily verified that $p(\phi) = \vec{I}_\phi \cdot \vec{p}$ agrees in value with the results for each of $A_1, A_2 \wedge A_3, \tau$ in Example 8.
Note the indicator vectors are just the columns of the truth table!

Three remarks are in order.

1. The vector representation gives us a geometric intuition for the space of probability functions on the language. Each probability function can be associated with a point on the convex hull of the simplex formed by the tips of the vectors representing the atomic states (which are simply the unit vectors along each axis).
2. We can draw a parallel between evaluations and probability functions. An evaluation is fully determined by the atomic propositions and a probability function is fully determined by the atomic states.
3. Theorem 13 tells us that the probability of a proposition is the sum of the probability of the propositions of the atomic states which entail the proposition. Since by Proposition 5 being entailed by an atomic state is the same as being true under one of the evaluations, the probability of a proposition can be thought of as a measure of the possibility that the proposition is true. A connection between probability and modal logic motivated by this observation is briefly explored in Appendix A.

Before moving on, we define what it means for propositions to be mutually exclusive which is *not* the same as them being disjoint. In addition, we define what we will call ψ-conditional probability distributions.

Definition 10. *Let $\mathcal{L}$ be a finite propositional language equipped with probability function p. We say two propositions ϕ, ψ are **mutually exclusive** exactly when*

$$p(\phi \wedge \psi) = 0.$$

Clearly, if two propositions are disjoint then they are mutually exclusive, but the other way does not always hold. Disjoint is a logical notion which says that the conjunction of two propositions is a contradiction and mutually exclusive is a probabilistic notion which says that the conjunction of two propositions has probability 0.

Definition 11. *A ψ-**conditional probability** function $p_\psi : \mathcal{F} \to [0,1]$ is a probability function which additionally satisfies*

$$p_\psi(\psi) = 1.$$

Note that Definition 11 is not the same as the usual definition of conditional probabilities which are defined in terms of an already existing probability function. Typically one defines a conditional probability function $p(\cdot \mid \psi)$ in terms of an unconditional probability function p by

$$p(\phi \mid \psi) := \frac{p(\phi \wedge \psi)}{p(\psi)}.$$

$p(\cdot \mid \psi)$ happens to be a ψ-conditional probability function because

$$p(\psi \mid \psi) = \frac{p(\psi \wedge \psi)}{p(\psi)} = \frac{p(\psi)}{p(\psi)} = 1.$$

We define ψ-conditional probability functions in order to specify a more restricted type of probability function which will come into play to show how a set-theoretic probability space can be represented in a propositional probability space (Section 4.2).

3.2.3 Standard Semantics

In this section we discuss how entailment can be generalized from propositional logic to probabilistic logic to give a concrete meaning to probabilistic inference. The account presented here comes from [5, Ch.7] with some modifications to naming conventions.

Recall from Definition 3 that in propositional logic

$$\phi_1, ..., \phi_k \models \psi$$

when for any evaluation v under which the premises $\phi_1, ..., \phi_k$ are true, v makes the conclusion ϕ true. As shown in Section 3.2, probability functions naturally generalize evaluation functions. Therefore, to generalize entailment, we will essentially just substitute a probability function in for the evaluation in Definition 3.

Definition 12. *Premises $\phi_1^{X_1}, ..., \phi_k^{X_k}$ **p-entails (probabilistic entailment)** conclusion ψ^Y, denoted*

$$\phi_1^{X_1}, \ldots, \phi_k^{X_k} \approx \psi^Y, \tag{16}$$

exactly when for all probability functions p:

$$\textit{if } p(\phi_1) \in X_1, \ldots, p(\phi_k) \in X_k \textit{ then } p(\psi) \in Y, \tag{17}$$

where $X_1, \ldots, X_k, Y \subseteq [0,1]$.

Definition 12 is referred to in [5, Ch. 7] as the *standard semantics*. Note that $X_1, \ldots, X_k, Y$ are subsets of $[0,1]$. We can think of $\phi_1^{X_1}, \ldots, \phi_k^{X_k}$ as constraints on the probability function of the form $p(\phi_i) \in X_i$ where $(i \in \{1, \ldots, k\})$ and then $p(\psi) \in Y$ as saying that the possible probabilities of ψ under those constraints are contained in Y. The next proposition shows how entailment is a special case of p-entailment.

Proposition 15. *If* $\phi_1^{\{1\}}, ..., \phi_k^{\{1\}} \approx \psi^{\{1\}}$ *then* $\phi_1, ..., \phi_k \models \psi$.

Proof. Suppose for an evaluation v, $v(\phi_1) = 1, \ldots, v(\phi_k) = 1$. Then since evaluations are also probability functions (Proposition 8), by Definition 12, $v(\psi) = 1$. □

Something to note is that if $Y = [0,1]$ then any p-entailment holds as we can see from the next proposition.

Proposition 16. $\phi_1^{X_1}, \ldots, \phi_k^{X_k} \approx \psi^{[0,1]}$ *for any* $\phi_1, ..., \phi_k, \psi$.

Proof. If p satisfies $\phi_1^{X_1}, \ldots, \phi_k^{X_k}$ then regardless of what p is, $p(\psi) \in [0,1]$. □

In propositional logic we often ask what propositions follow from certain other propositions. In probabilistic logic we can ask questions of the form

$$\phi_1^{X_1}, \ldots, \phi_k^{X_k} \approx \psi^?$$

which asks for the set of possible probabilities for our conclusion, Y, given some premises. By Proposition 16 we can always replace ? with $[0,1]$ which is not very useful. What we mean by a question like

$$\phi_1^{X_1}, \ldots, \phi_k^{X_k} \approx \psi^?$$

is for the smallest possible set for which the entailment holds. This gives us the set of possible beliefs to which that the premises absolutely constrain the conclusion. We accordingly define minimal p-entailment.

Definition 13. *Premises* $\phi_1^{X_1}, ..., \phi_k^{X_k}$ ***minimally p-entail*** *conclusion* ψ^Y*, exactly when* $\phi_1^{X_1}, \ldots, \phi_k^{X_k} \approx \psi^Y$ *and there is no subset* $Z \subseteq Y$ *such that* $\phi_1^{X_1}, \ldots, \phi_k^{X_k} \approx \psi^Z$.

Example 10. *The following are some instances of some minimal p-entailments.*

1. $A_1^{[0.2,0.7]} \approx (\neg A_1)^{[0.3,0.8]}$
2. $A_1^{\{0.5\}}, A_2^{[0,0.6]} \approx (A_1 \wedge A_2)^{[0,0.5]}$
3. $(A_1 \wedge A_2)^{\{0.25\}}, (A_1 \vee \neg A_2)^{\{1\}} \approx A_2^{\{0.25\}}$.

p-entailment show us how we can generalize the notion of inference from propositional logic to a probabilistic logic.

4 Comparing the frameworks

Having developed the finite propositional framework for probability, we now compare it to the finite set-theoretic framework for probability theory. The main argument we make in favor of the propositional framework is that although both frameworks can abstractly represent one another, the elementary outcomes in the set-theoretic framework are conceptually reducible to the propositions in a propositional framework in a way that the reverse does not hold.

To make the argument concrete we will consider two cases in which we draw from a bag of three balls. We will analyze the probability space of both frameworks for each case.

Case 1: When we draw from the bag, we might come out empty-handed or we might pull out several balls at once (they are tiny enough like marbles to do so).

Case 2: When we draw from the bag, we draw one of the three balls.

To begin let us make explicit what a probability space in each framework is. Definition 14 draws from [8, Ch. 1] and Definition 15 is original.

Definition 14. *A **set-theoretic probability space** (Ω, Σ, p) consists of a set of elementary outcomes*

$$\Omega = \{e_1, ..., e_n\},$$

a sigma algebra of elements from Ω

$$\Sigma,$$

and a probability function

$$p,$$

defined by Kolmogorov's axioms (1), (2).

Definition 15. *A **propositional probability space** $(Pr, \mathcal{F}, p)$ consists of a set of atomic propositions*

$$Pr = \{A_1, ..., A_n\},$$

a set of all propositions generated from Pr (Definition 1)

$$\mathcal{F},$$

and a probability function

$$p,$$

*defined by the structure preserving axioms **A1**, **A2**, **A3**.*

4.1 Case 1

4.1.1 Set-theoretic

To represent case 1 in the set-theoretic framework, we must first identify the elementary outcomes. There are 8 elementary outcomes that could arise from the experiment of drawing balls. These outcomes correspond to which combination of balls were drawn. Label the balls b1, b2, b3. The elementary outcomes would be

$$e_0 : \text{no ball drawn}$$
$$e_1 : \text{only b1 drawn}, \; e_2 : \text{only b2 drawn}, \; e_3 : \text{only b3 drawn}$$
$$e_4 : \text{only b1, b2 drawn}, \; e_5 : \text{only b2, b3 drawn}, \; e_6 : \text{only b1, b3 drawn}$$
$$e_7 : \text{only b1, b2, b3 drawn}.$$

The elementary outcome set Ω would be $\{e_1, \dots e_8\}$ and the set of events $\mathcal{P}(\Omega)$. The probaility space would therefore be

$$(\{e_1, \dots e_8\}, \mathcal{P}(\Omega), p)$$

for some probability function p.

4.1.2 Propositional

To represent case 1 in the propositional framework, we must identify the basic atoms of the phenomena that we can describe (this is different from the elementary outcomes of the experiment). There are 3 atomic propositions A_1, A_2, A_3 corresponding to what balls were drawn. A_i stands in for the proposition 'ball i was drawn.'. Therefore the probability space would be

$$(\{A_1.A_2, A_3\}, \mathcal{F}, p)$$

for some probability function p.

Note that the atomic states correspond to all the possibilities we might observe. For example,

- $A_1 \wedge A_2 \wedge A_3$ says all balls will be in hand after draw.
- $A_1 \wedge \neg A_2 \wedge A_3$ says ball 1 and ball 3 will be in hand after draw.
- $\neg A_1 \wedge \neg A_2 \wedge \neg A_3$ says no ball is in hand after draw.

4.1.3 Comparison

The most striking thing to observe is that in the set-theoretic framework we needed 8 elementary outcomes whereas in the propositional framework we only needed 3 atomic propositions to fully characterize the scenario. We can learn from this that the elementary outcomes of an experiment are not always the most elementary components we can break a problem into. Sometimes atomic propositions are a more fundamental unit of analaysis.

The above analysis also shows us how we can represent the propositional framework in a set-theoretic framework. If we label the 8 possible atomic states $\omega_1, \dots, \omega_8$ then we can take Ω, now the set of atomic states, to be the set of elementary outcomes in the set-theoretic framework. For example, $\neg A_1 \wedge \neg A_2 \wedge \neg A_3$ would correspond to e_0 since both express that no ball was drawn. A proposition like $A_1 \wedge A_2$ can be represented by what is effectively its disjunctive normal form $(A_1 \wedge A_2 \wedge A_3) \vee (A_1 \wedge A_2 \wedge \neg A_3)$ as the event $\{A_1 \wedge A_2 \wedge A_3, A_1 \wedge A_2 \wedge \neg A_3\}$ which would correspond to $\{e_8, e_4\}$. By Theorem 13, The respective probability function acts the same on both.

$$p(A_1 \wedge A_2) = p(A_1 \wedge A_2 \wedge A_3) + p(A_1 \wedge A_2 \wedge \neg A_3)$$

and by (2),

$$\begin{aligned} p(\{\{e_8, e_4\}\}) &= p(\{\{e_8\}\} \cup \{e_4\}) \\ &= p(\{e_8\}) + p(\{e_4\}). \end{aligned}$$

Given that we can convert a propositional probability space to a set-theoretic probability space as outlined above, wouldn't that mean that a set-theoretic framework is at least as powerful as a propositional framework, so that there is no reason for preferring a propositional framework? While we *can* wrap atomic states in sets to represent a propositional analysis in a set-theoretic analysis, to do so is undesirable for two reasons. First, the wrapping things in sets is unnecessary when you can work directly with atomic states. Second, the atomic states are not the most fundamental unit of conceptual analysis. To treat atomic states as elementary outcomes is to ignore what is conceptually more primitive, the atomic propositions. In a meaningful sense, the set-theoretic representation is, in this case, conceptually reducible to the propositional framework since the elementary outcomes can really be thought of as being composed of a conjunction of atomic propositions.

4.2 Case 2

4.2.1 Set-theoretic

In case 2, we are told that we draw one of the three balls so there are three elementary outcomes of the experiment:

$$e_1 : \text{only b1 drawn}, e_2 : \text{only b2 drawn}, e_3 : \text{only b3 drawn}.$$

The set-theoretic probability space would be

$$(\{e_1, e_2, e_3\}, \mathcal{P}(\{e_1, e_2, e_3\}), p).$$

4.2.2 Propositional

The propositional account of this case is slightly more nuanced. Similar to case one, we will label the elementary observations with atomic states A_1, A_2, A_3 in which A_i stands in for the proposition 'ball i was drawn.' A_1, A_2, A_3 are merely symbols with no intrinsic logical connection between them. The scenario, however, does specify a logic relation between A_1, A_2, A_3 because it says that exactly one of them will be drawn. It is our responsibility to formalize how A_1, A_2, A_3 are logically related to one another. The assumption that exactly one of the three balls will be drawn is more precisely expressed by the probabilistic assumption that A_1, A_2, A_3 are exhaustive and pairwise mutually exclusive.

$$p(A_1 \vee A_2 \vee A_3) = 1, \tag{18}$$
$$p(A_i \wedge A_j) = 0 \text{ for } i \neq j. \tag{19}$$

We can think of these assumptions as constituting a logic model of our phenomena. They are the axioms of our phenomena in a similar way to how the Peano axioms pick out particular intuitive properties of the phenomena we call natural numbers from which we can derive more facts about numbers. 18 and 19 are the axioms of our phenomena from which we can deduce, for example, that the probability of drawing ball 1 or 2 is the sum of their respective individual probabilities:

$$p(A_1 \vee A_2) = p(A_1) + p(A_2) - p(A_1 \wedge A_2) = p(A_1) + p(A_2).$$

All this to say, case 2 is represented by a propositional probability space

$$(\{A_1, A_2, A_3\}, \mathcal{F}, p_\psi)$$

where $\mathcal{F}$ is generated from $\{A_1, A_2, A_3\}$ according to Definition 1 and p_ψ is a ψ-conditional probability function (Definition 11) where

$$\psi := (A_1 \vee A_2 \vee A_3) \wedge \neg(p(A_1 \wedge A_2)) \wedge \neg(p(A_2 \wedge A_3)) \wedge \neg(p(A_1 \wedge A_3)).$$

p_ψ builds the assumptions 18 and 19 into the probability space.

4.2.3 Comparison

We can see from these examples how any set-theoretic probability space can be converted to a propositional probability space. Given a set-theoretic probability space $(\{e_1, \dots, e_n\}, \Sigma, p)$ we can convert it to a propositional probability space $(Pr, \mathcal{F}, p_\psi)$ by setting $Pr = \{e_1, \dots, e_n\}$, and choosing p_ψ to be the ψ-conditional probability function such that

$$p_\psi(e_{i_1} \vee \ldots \vee e_{i_k}) := p(\{e_{i_1}, \ldots e_{i_k}\})$$

where

$$\psi := [\bigwedge_{\substack{i,j \in \{1,\ldots,n\} \\ i \neq j}} \neg(e_i \wedge e_j)] \wedge (e_1 \vee ... \vee e_n).$$

Aside from showing how we can represent set-theoretic probability spaces as propositional probability spaces, we can learn from this conversion process that the elementary outcomes of an experiment are essentially just mutually exclusive and exhaustive propositions. From this point of view, once again, propositions are the conceptually more fundamental unit of analysis and elementary outcomes are taken to be special kinds of propositions. In case one, we saw that the elementary outcomes could be understood as atomic states and in case two the elementary outcomes are simply mutually exclusive and exhaustive propositions! In both cases the notion of elementary outcomes is conceptually reducible to a propositional perspective!

This is perhaps not so surprising when we observe that the number of propositions generated from n atomic propositions is 2^{2^n} (there are 2^n atomic states and each proposition is some disjunction of atomic states by Theorem 6) while the maximum number of events that can be generated from an outcome space of n elementary outcomes is 2^n (by taking the power set). A propositional framework is simply able to express more using less primitives.

Although a propositional framework is conceptually more fundamental than a set-theoretic one, from the vantage point of propositions, we can understand when a set-theoretic probability space might have some utility. By wrapping atomic propositions in sets to represent the outcomes of an experiment one is effectively using a convenient data structure that builds in the assumption that the atomic propositions are mutually exclusive. If someone knows that their atomic propositions are mutually exclusive, to wrap them in sets might have computational advantages.

5 Summary

We have shown how we can naturally recover a finitely additive probability function on propositions by aiming to extend evaluation functions from the discrete truth values of $\{0, 1\}$ to the interval $[0, 1]$ in a way that preserves key invariants. This differs from the traditional approach to probability theory on sets of events which uses the Kolmogorov axioms to define probability functions because they give the right "formal properties" [2]. We further showed how we can generalize the notion of entailment in propositional logic to a probabilistic logic using the notion of p-entailment. This enables us to combine the conceptual infrastructure of logical inference with probabilities (which has applications to systematizing inductive logics based on probabilities [3]).

After setting up both frameworks, we showed that both the set-theoretic and the propositional frameworks can represent one another (we can convert one to the other). However, despite the frameworks being representationally equivalent, we also showed that elementary outcomes in a set-theoretic probability space can be more fundamentally understood as special kinds of propositions.

Given that probability functions on propositions are naturally motivated as extending propositional logic and elementary outcomes can be conceptually reduced to special kinds of propositions, we conclude that finite probability theory should be done and taught in a propositional framework.

Appendix A Modal Logic and Probability

In this section we will examine a result connecting the logic of possibility (modal logic) with probability. Some background in modal logic is assumed because this is only intended as a brief section to note an interesting result, the meaning of which I have not yet figured out. Initially I attempted to use the result to show that we can think of degree of belief in a proposition as a measure of the possibility that the proposition is true in the real world, but I was not able to develop a sufficiently compelling account.

In modal logic we extend our language to include statements of the form:

- $\phi \to \psi$ which is usually read 'if ϕ then ψ' and is semantically equivalent to $\neg\phi \vee \psi$.
- $\Box\phi$ which is usually read 'ϕ is necessary'.
- $\Diamond\phi$ which is usually read 'ϕ is possible'.

We will only consider a fragment of the modal logic $\boldsymbol{K}$ where $\Box$ and $\Diamond$ can not scope over other $\Box$'s and $\Diamond$'s. The axioms for $\boldsymbol{K}$ [4] can be stated for $\phi, \psi \in \mathcal{F}$

N If $\models \phi$ then $\Box\phi$.

D $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$.

We now show that for a probability distribution p, the statement $p(\phi) = 1$ satisfies **N** and **D** (over the fragment we are considering) when it is substituted in for $\Box\phi$.

N If $\models \phi$ then by Definition 4 ϕ is a tautology so by ***A1***, $p(\phi) = 1$. This shows that

$$\text{If } \models \phi \text{ then } p(\phi) = 1.$$

D Suppose $p(\phi \to \psi) = 1$ this is the same as saying $p(\neg\phi \vee \psi) = 1$. Therefore, by ***A2***, $p(\neg\phi) + p(\psi) - p(\neg\phi \wedge \psi) = 1$. If $p(\phi) = 1$ then $p(\neg\phi) = 0$ (by ***A1***) so $p(\psi) - p(\neg\phi \wedge \psi) = 1$. Since $p(\psi) \leq 1$ and $p(\neg\phi \wedge \psi) \geq 0$, $p(\psi) = 1$. This shows that

$$[p(\phi \to \psi) = 1] \to ([p(\phi) = 1] \to [p(\psi) = 1]).$$

Given that $p(\phi) = 1$ can stand in for $\Box\phi$, what would stand in for $\Diamond\phi$? In modal logic,

$$\neg\Box\phi \equiv \Diamond\neg\phi.$$

We observe a similar relationship in probability:

$$p(\phi) \neq 1 \Leftrightarrow p(\phi) < 1 \Leftrightarrow p(\neg\phi) > 0$$

indicating that we should take $p(\phi) > 0$ to stand for $\Diamond\phi$.

Given that $p(\phi) = 1$ obeys the axioms for a fragment of ***K*** might we be able to interpret probability or belief modally?

References

[1] E. T. Jaynes,*Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, 2003.

[2] H. Bauer,*Probability Theory*, De Gruyter, Berlin, 1996.

[3] H. Rolf, J. Romeijn, G. Wheeler, J. Williamson, *Probabilistic Logics and Probabilistic Networks*, Springer, Dordrecht, 2011.

[4] J. Garson, "Modal Logic", The Stanford Encyclopedia of Philosophy (2023), URL = ¡https://plato.stanford.edu/archives/spr2023/entries/logic-modal¿.

[5] J. Williamson, *In Defense of Objective Bayesianism*, Oxford University Press, Oxford, 2010.

[6] K. S. Van Horn,: Constructing a Logic of Plausible Inference: A Guide to Cox's Theorem, *International Journal of Approximate Reasoning*, 34(1), 3-24 (2003).

[7] M. G. Titelbaum, *Fundamentals of Bayesian Epistemology*, Oxford University Press, New York, 2022.

[8] M. M. Rao, R. J. Swift, *Probability Theory with Applications*, Springer, New York, 2006.