Probability Functions in the Context of Signed Involutive Meadows (Extended Abstract)

Jan A. Bergstra and Alban Ponse(✉)

Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
A.Ponse@uva.nl
https://staff.science.uva.nl/j.a.bergstra/,
https://staff.science.uva.nl/a.ponse/

Abstract. The Kolmogorov axioms for probability functions are placed in the context of signed meadows. A completeness theorem is stated and proven for the resulting equational theory of probability calculus. Elementary definitions of probability theory are restated in this framework.

Keywords: Meadow · Bayes' theorem · Bayesian reasoning

1 Introduction

The Kolmogorov axioms for probability functions may be considered a module that can be included in a variety of more or less formalized contexts. We will propose and investigate some consequences of these axioms when placed in the context of involutive meadows, that is meadows where inverse is an involution following the terminology of [7].

In particular we will discuss an axiomatization of a probability function (PF) on a Boolean algebra. The Boolean algebra serves as an event space, the PF defined on it produces elements of (values in) a signed meadow that serve as probabilities. Special focus is on the case where values are chosen in the signed meadow of real numbers. The following objectives motivate the line of development in this paper.

1. To develop an approach towards strictly equational reasoning about probability.
2. To provide a finite loose equational specification of probability functions.
3. To provide a useful completeness result for equational axioms of probability functions.
4. To investigate some total versions of the conditional probability operator.
5. To initiate the development of an application for the theory of signed meadows as outlined in [4,5].

This paper is a shortened version of http://arxiv.org/abs/1307.5173v4.

P. James and M. Roggenbach (Eds.): WADT 2016, LNCS 10644, pp. 73–87, 2017.
https://doi.org/10.1007/978-3-319-72044-9_6

We will produce an axiom system consisting of twenty-six equational axioms covering Boolean algebra, meadows, the sign function, and the PF. Then we will introduce several derived operators and prove a number of simple facts, including Bayes' theorem.

These axioms constitute a finite equational basis for the class of Boolean algebra based, real-valued PFs. In other words, the completeness results of [4,5] extend to the case with Boolean algebra based PFs. We understand this result to convey that the set of twenty-six axioms is complete in a reasonable sense.

The paper is structured as follows: in the remainder of this section we discuss the concept of a meadow in more detail and provide a survey of relevant design options. In Sect. 2 we introduce some preliminaries. In Sect. 3 we provide equational axioms for a PF, and in Sect. 4 we discuss completeness. In Sect. 5 we consider multi-dimensional probability functions, and Sect. 6 contains some concluding remarks.

1.1 A Survey of Design Options for the Inverse of 0

A meadow is a ring-like structure equipped with an inverse function. A ring based meadow expands a ring with a one place inverse function (inversive notation), or a two place division function (divisive notation). The terms 'inversive notation' and 'divisive notation' were coined in [6].

The key design choice that needs to be made when contemplating a meadow concerns the way it handles the inverse of 0. In a rather scattered literature on the subject a plurality of different options has been developed and studied, though in varying levels of detail. A brief survey of these endeavours sets the stage for the plan of this paper. The listing below is incomplete, but it contains all proposals for which we have been able to find an unambiguous description. As a criterion regarding this judgement we have required that (i) it must be possible to find out when a closed expressions written using $0, 1, +, -, \cdot, (-)^{-1}$ is considered to have a value in the mathematical structure at hand, (ii) for two closed expressions both having a value it must be possible to determine equality in the same structure, and (iii) the relation between inverse and division must be transparent. We will distinguish three design options for ring based meadows and three design options for non-ring based meadows. We will first survey design options for non-ring based meadows.

Non-ring Based Meadows

Three options for setting the inverse of zero in a non-ring based meadow stand out, each involving an error value which fails to meet the requirements of a ring. Distinguishing these options is facilitated by making use of a uniform terminology.

Natural inverse. If 0^{-1} is equated with an unsigned infinite value, often denoted by ∞, then 0 is said to have a natural inverse. The use of natural inverse in mathematics dates back to Riemann at least. Wheels are the prominent instance of meadows with natural inverse, see [9].

Signed natural inverse. If the inverse of zero is equated with a signed infinite value (written say as $+\infty = \infty$, which differs from $-\infty$) we propose to speak of a signed natural inverse. This design choice underlies the transreals and transrationals, see [16].

Common inverse. If the inverse of zero is equated with an error value $\mathbf{a}$ then, following [8], zero is said to have a common inverse. Common meadows are meadows based on common inverse. The error value $\mathbf{a}$ satisfies $x + \mathbf{a} = x \cdot \mathbf{a} = -\mathbf{a} = \mathbf{a}$ and for that reason fails to comply with the requirements for a ring $(0 \cdot x = 0)$. Moreover, the error value is unique.

Also in the case of natural inverse and signed natural inverse, the error value(s) fail to comply with the requirements for a ring $(0 \cdot x = 0)$.

Ring Based Meadows
For ring based meadows three options may be distinguished.

Partial inverse. The most prominent ring based meadow leaves the inverse of 0 undefined and considers inverse to be a partial function.
Working with partial inverse deviates from mathematical practice to the extent that questions like whether or not $1/0 = 2/0$ must be taken seriously. When dealing with partial inverse there are no semantic questions about it, but the choice of a logic of partial functions leaves substantial room for design variation, beginning with a choice between three ways of looking at the truth value of say $1/0 = 1/0$: is it considered as being true, or as being false in an overarching two-valued logic, or as not being true in an overarching logic which is not two-valued.

Symmetric inverse. If the meadow is based on a regular ring and the value of 0^{-1} is taken to be 0, 0 is said to have a symmetric inverse. The meadows of [4,5] and several preceding papers are ring based meadows with symmetric inverse. Alternatively this case is referred to as featuring an involutive inverse, and such meadows are referred to as involutive meadows.

Non-involutive inverse. If the inverse of 0^{-1} is taken to be different from 0, $(x^{-1})^{-1} = x$ cannot hold, that is inverse is not an involution, and inverse is said to be non-involutive. The non-involutive meadows discussed in [7] that satisfy $0^{-1} = 1$ are ring based meadows with an asymmetric inverse. If the inverse of 0^{-1} is taken to be say 17 or any (rational or real) number different from 0 and 1, 0 is said to have an ad hoc non-involutive inverse. Ad hoc non-involutive inverses come into play when formalizing the theory of fields in first order logic in the presence of a function symbol for either inverse or division (or both).

1.2 Working with Involutive Ring Based Meadows

In this paper we will work exclusively with ring based involutive meadows, which will be referred to simply as meadows. The motivation for this choice is that it appears to be a most straightforward way to pursue the objectives that were listed above. However, we do not claim that for the purpose of developing an equational approach to probability working with ring based meadows is the best option, neither do we claim that among the three options for ring based meadows working with a symmetric inverse is best suited to this objective.

2 Boolean Algebras and Meadows

In this section we specify the mathematical context on which our axiomatization is based. In particular, we provide specifications for Boolean algebras (Sect. 2.2), and for events and (signed) meadows (Sect. 2.3).

2.1 Boolean Algebras

A Boolean algebra $(B, +, \cdot, ', 1, 0)$ may be defined as a system with at least two elements such that $\forall x, y, z \in B$ the well-know postulates of Boolean algebra are valid. Because we want to avoid overlap with the operations of a meadow, we will consider Boolean algebras with notation from propositional logic, thus consider $(B, \vee, \wedge, \neg, \top, \bot)$ and adopt the axioms in Table 1. In [14] it was shown that the axioms in Table 1 constitute an equational basis.

Table 1. *BA*, a self-dual equational basis for Boolean algebras

$(x \vee y) \wedge y = y$	(1)	$x \vee (y \wedge z) = (y \vee x) \wedge (z \vee x)$	(4)
$(x \wedge y) \vee y = y$	(2)	$x \wedge \neg x = \bot$	(5)
$x \wedge (y \vee z) = (y \wedge x) \vee (z \wedge x)$	(3)	$x \vee \neg x = \top$	(6)

2.2 Valuated Boolean Algebras and Some Naming Conventions

A Boolean algebra can be equipped with a valuation v that assigns to its elements values in a signed meadow.

In this paper we will investigate the special case where the valuation function of a valuated Boolean algebra is a probability function by requiring that the valuation satisfies the Kolmogorov axioms for probability functions cast to the setting of signed meadows.

By way of notational convention we will from now on assume that E (for events) is the name of the carrier of a Boolean algebra, and that V (for values) names the carrier of the meadow in a valuated Boolean algebra.

2.3 Events and Signed Meadows

The set of axioms in Table 2 specifies the class of meadows. In the setting of probability functions the elements of the underlying Boolean algebra are referred to as events.[1] We will use "value" to refer to an element of a meadow,[2] and a probability function is a valuation (from events to the values in a signed meadow).[3]

Table 2. *Md*, a set of axioms for meadows

$(x+y)+z = x+(y+z)$	(7)	$x \cdot y = y \cdot x$	(12)
$x+y = y+x$	(8)	$1 \cdot x = x$	(13)
$x+0 = x$	(9)	$x \cdot (y+z) = x \cdot y + x \cdot z$	(14)
$x+(-x) = 0$	(10)	$(x^{-1})^{-1} = x$	(15)
$(x \cdot y) \cdot z = x \cdot (y \cdot z)$	(11)	$x \cdot (x \cdot x^{-1}) = x$	(16)

An expression of type E is an event expression or an event term, an expression of type V is a value expression or equivalently a value term. In the signature of a valuated Boolean algebra there is just one notation for a probability function, the function symbol P.[4]

In a meadow equipped with an ordering $<$, the sign function $\mathbf{s}(\_)$ is defined by

$$\mathbf{s}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } 0 < x. \end{cases}$$

The axioms in Table 3 specify the sign function in an equational manner. Before commenting on these axioms, we define the conditional $p \lhd q \rhd r$ expressing a form of "if-then-else", notations for a division operator, absolute value, and orderings. Furthermore, we adopt the convention to write $x - y$ for $x + (-y)$. Let

Table 3. *Sign*, a set of axioms for the sign operator

$\mathbf{s}(1_x) = 1_x$	(17)	$\mathbf{s}(x^{-1}) = \mathbf{s}(x)$	(20)
$\mathbf{s}(0_x) = 0_x$	(18)	$\mathbf{s}(x \cdot y) = \mathbf{s}(x) \cdot \mathbf{s}(y)$	(21)
$\mathbf{s}(-1) = -1$	(19)	$0_{\mathbf{s}(x)-\mathbf{s}(y)} \cdot (\mathbf{s}(x+y) - \mathbf{s}(x)) = 0$	(22)

[1] Events are closed under $- \vee -$, which represents alternative occurrence and $- \wedge -$, which represents simultaneous occurrence, and under negation.

[2] Rational numbers and real numbers are instances of values.

[3] We will exclude probability functions with negative values, a phenomenon known in non-commutative probability theory, leaving the exploration of that kind of generalization to future work.

[4] In some cases the restriction to a single probability function P is impractical and providing a dedicated sort for such functions brings more flexibility and expressive power. This expansion may be achieved in different ways.

p, q and r range over V, the carrier of the signed meadow in a valuated Boolean algebra, then

$$1_p =_{\texttt{def}} p \cdot p^{-1}, \qquad p/q =_{\texttt{def}} \frac{p}{q},$$
$$0_p =_{\texttt{def}} 1 - 1_p, \qquad |p| =_{\texttt{def}} \mathbf{s}(p) \cdot p,$$
$$p \triangleleft q \triangleright r =_{\texttt{def}} 1_q \cdot p + 0_q \cdot r, \qquad p < q =_{\texttt{def}} \mathbf{s}(q-p) = 1,$$
$$\frac{p}{q} =_{\texttt{def}} p \cdot q^{-1}, \qquad p \leq q =_{\texttt{def}} \mathbf{s}(\mathbf{s}(q-p)+1) = 1.$$

In Table 3, axiom (22) is an equational representation of the conditional equation $\mathbf{s}(x) = \mathbf{s}(y) \;\rightarrow\; \mathbf{s}(x+y) = \mathbf{s}(x)$. Finally, the equivalences

$$p \geq 0 \iff \mathbf{s}(\mathbf{s}(p)+1) = 1 \iff p = \mathbf{s}(p) \cdot p = |p|$$

are provable from $Md + Sign$ (this follows from Theorem 4.1.1 below).

We will also consider the subclass of signed *cancellation meadows*. A cancellation meadow satisfies the Inverse Law (IL) of Table 4.

Table 4. Inverse law (IL)

$x \neq 0 \;\rightarrow\; x \cdot x^{-1} = 1$

3 Signed Meadow Based Probability Calculus

In Sect. 3.1 we formulate axioms for a probability function. Following the methods of abstract data type specification we will focus on axioms in equational form. Then, we discuss a plurality of versions of the conditional probability operator (Sect. 3.2) and some properties thereof, in particular versions of Bayes' theorem. Finally, we consider independent events (Sect. 3.3).

3.1 Equational Axioms for a Probability Function

In Table 5 we define the set PF_P of axioms for a probability function. These axioms represent Kolmogorov's axioms in the context of a Boolean algebra (rather than a universe of sets) and a signed meadow (instead of a field). Axiom (25) expresses that the sign of $P(x)$ is nonnegative. Axiom (26) distributes P over finite unions. In the absence of an infinitary version of axiom (26) we consider these axioms to constitute an axiomatization for the restricted concept of probability functions only, rather than for probability measures in general.

In combination with the axioms $BA + Md$, the two axioms (24) and (26) in Table 5 can be replaced by the single axiom

$$P(x) = P(x \wedge y) + P(x \wedge \neg y) \qquad (\dagger)$$

Table 5. PF_P, a set of axioms for a probability function with name P

$P(\top) = 1$	(23)	$P(x) = \lvert P(x) \rvert$	(25)
$P(\bot) = 0$	(24)	$P(x \vee y) = P(x) + P(y) - P(x \wedge y)$	(26)

where the expressions $x \wedge y$ and $x \wedge \neg y$ characterize two disjoint (mutually exclusive) events: axiom (24) follows from $P(x) = P(x \wedge x) + P(x \wedge \neg x)$, thus $P(x \wedge \neg x) = P(\bot) = 0$, and axiom (26) follows from

$$P(x \vee y) \overset{\dagger}{=} P((x \vee y) \wedge x) + P((x \vee y) \wedge \neg x) = P(x) + P(y \wedge \neg x)$$

and $P(y) \overset{\dagger}{=} P(y \wedge x) + P(y \wedge \neg x)$, thus $P(y \wedge \neg x) = P(y) - P(x \wedge y)$. Conversely, axiom (†) follows from (24) and (26):

$$P(x) = P((x \wedge y) \vee (x \wedge \neg y)) = P(x \wedge y) + P(x \wedge \neg y) - P(\bot). \quad (\ddagger)$$

Theorem 3.1.1 (Disjoint event factorization). *$BA + Md + PF_P \vdash P(x) = P(x \wedge y) + P(x \wedge \neg y)$.*

Proof. This is (‡), which is shown above. □

Theorem 3.1.2 (Probability upper bound). *$BA + Md + Sign + PF_P \vdash P(x) \leq 1$.*

Proof. First notice $1 = P(\top) = P(x \vee \neg x) = (P(x) + P(\neg x)) - P(\bot) = P(x) + P(\neg x)$, so $P(x) = 1 - P(\neg x)$. Because $P(\neg x) \geq 0$ we conclude $P(x) \leq 1$. □

The following theorem asserts in equational form the conditional equation $P(y) = 0 \rightarrow P(x \wedge y) = 0$, using inversive notation.

Theorem 3.1.3 *$BA + Md + Sign + PF_P \vdash P(x \wedge y) \cdot P(y) \cdot P(y)^{-1} = P(x \wedge y)$.*

Proof. With help of Theorem 4.1.1, see there. □

3.2 Conditional Probability as a Total Operator: Four Options

Conditional probability $P(x \mid y)$ of event x relative to event y is conventionally understood as a partial function of x and y, defined only if $P(y)$ is nonzero. The objective of developing an equational logic for probability theory suggests that total versions of the conditional probability operator ought to be contemplated.

Conditional probability defined according to Kolmogorov is written below as $P^\star(x \mid y)$, where variables x and y range over E, and is defined by

$$P^\star(x \mid y) =_{\texttt{def}} \frac{P(x \wedge y)}{P(y)} \triangleleft P(y) \triangleright \uparrow.$$

Here $\uparrow$ denotes that the result is undefined.[5] The key advantage of partial conditional probability is that one does not introduce a value for, say $P(x \mid \bot)$ which might be subsequently disputed. Four ways of making conditional probability defined on all inputs will now be distinguished.

Definition 3.2.1 (Zero-totalized conditional probability).

$$P^0(x \mid y) =_{\texttt{def}} \frac{P(x \wedge y)}{P(y)}.$$

We notice that $P^0(\top \mid \bot) = 0$, a choice for which no convincing philosophical motivation can be put forward. Two advantages can be put forward in favour of $P^0(-|-)$: the logical simplicity that comes with it being total and the calculational simplicity that comes with choosing 0 as a value for $P^0(x \mid y)$ when $P(y) = 0$. The following properties are immediate:

$$P^0(x \mid x) = \frac{P(x)}{P(x)} \quad \text{and} \quad P(x) = P(x) \cdot P^0(x \mid x).$$

Moreover we have 'joint probability factorization':

$$P(x \wedge y) = P(x \wedge y) \cdot P(y) \cdot P(y)^{-1} = (P(x \wedge y)/P(y)) \cdot P(y) = P^0(x \mid y) \cdot P(y),$$

and 'total probability':

$$\begin{aligned} P(x) &= P(x \wedge y) + P(x \wedge \neg y) \\ &= P(x \wedge y) \cdot P(y) \cdot P(y)^{-1} + P(x \wedge \neg y) \cdot P(\neg y) \cdot P(\neg y)^{-1} \\ &= P^0(x \mid y) \cdot P(y) + P^0(x \mid \neg y) \cdot P(\neg y). \end{aligned}$$

Another illustration of the latter advantage is the derivability of Bayes' theorem in its simplest form (see Theorem 3.2.5.1).

Definition 3.2.2 (One-totalized conditional probability).

$$P^1(x \mid y) =_{\texttt{def}} \frac{P(x \wedge y)}{P(y)} \triangleleft P(y) \triangleright 1.$$

[5] We assume that in a context of partial functions an identity $t = r$ is valid if either both sides are undefined or both sides are defined and equal. This convention, however, leaves room for alternative readings of the expressions at hand. In particular the definition given for $x \triangleleft y \triangleright z$ implies that whenever t is undefined, so is $t \triangleleft r \triangleright s$. That is not a very plausible feature of the conditional and in the presence of partial operations the conditional operator requires a different definition. These complications are to some extent avoided, or rather made entirely explicit, when working with total functions. The use of the notation $P^\star(-|-)$ instead of the common notation $P(-|-)$ is justified by the fact that unavoidably $P^\star(-|-)$ inherits properties from the equational specification of the functions from which it has been made up. Such properties need not not coincide with what is expected from $P(-|-)$.

We will write $x \to y$ for $\neg x \vee y$. The principal advantage of one-totalized conditional probability over zero-totalized conditional probability is the validity of the following rule, which provides some intrinsic motivation for this design of conditional probability:

$$(x \to y) = \top \;\Rightarrow\; P^1(x \mid y) = 1.$$

If $\alpha \in \{\star, 0, 1\}$ then the function $P \circ^\alpha y =_{\mathtt{def}} \lambda x \in E.P^\alpha(x|y)$ is not a probability function for each y. In particular, if $P(y) = 0$, $P \circ^\alpha y$ will fail to comply with either $P \circ^\alpha y(\top) = 1$ or with $P \circ^\alpha y(\bot) = 0$. Now $\lambda P \in PF.P \circ^\alpha y$ being the well-known update operator that goes with some applications of Bayes' theorem, it is a reasonable requirement that this very operator becomes total as well. We will introduce two options for conditionalization which achieve this requirement.

Definition 3.2.3 (Safe conditional probability).

$$P^s(x \mid y) =_{\mathtt{def}} \frac{P(x \wedge y)}{P(y)} \lhd P(y) \rhd P(x).$$

We find that $P \circ^s y = P$ if $P(y) = 0$, which allows the view that $\lambda P.P \circ^s y$ is an operator mapping probability functions to probability functions for all events y, or stated differently that $\lambda y.(\lambda P.P \circ^s y)$ is a total mapping from events to probability function transformations. $P\circ^s$ is safe because it enforces no update when an inconsistency is observed.

Yet another way to achieve this property of a conditional update is to return an exceptional value, in this case the canonical probability function for an atomic event. An atom in E is an event $a \in E$ which satisfies $\mathtt{atom}(a) =_{\mathtt{def}} \forall x \in E.(x \wedge a = a \;\mathtt{OR}\; x \wedge a = \bot)$. For an atom $a \in E$ the probability function $\mathtt{pf}_a$ is defined by:

$$\mathtt{pf}_a(x) =_{\mathtt{def}} \begin{cases} 1 & \text{if}\, x \wedge a = a, \\ 0 & \text{if}\, x \wedge a = \bot. \end{cases}$$

Definition 3.2.4 (Exception raising conditional probability for atom $a \in E$).

$$P^{e/a}(x \mid y) =_{\mathtt{def}} \frac{P(x \wedge y)}{P(y)} \lhd P(y) \rhd \mathtt{pf}_a(x).$$

For $P^0(-|-), P^s(-|-)$, and $P^{e/a}(-|-)$ we are not aware of earlier definitions, whereas $P^1(-|-)$ has been considered by Adams in [1], and in subsequent literature. For a survey of conditional logic and conditional probabilities we refer to [12].

Of particular importance given its ubiquitous use is Bayes' theorem. Bayes' theorem takes different forms for different versions of conditional probability, and in each of these cases it appears as a consequence of $BA + Md + Sign + PF_P$.

Theorem 3.2.5 (Versions of Bayes' theorem). *In $BA + Md + Sign + PF_P$ the following equations are derivable:*

1. $P^0(x \mid y) = \dfrac{P^0(y \mid x) \cdot P(x)}{P(y)}$ $\quad (Bayes' \, theorem \, for P^0(-|-))$,
2. $P^1(x \mid y) = \dfrac{P^1(y \mid x) \cdot P(x)}{P(y)} \triangleleft P(y) \triangleright 1$ $\quad (Bayes' \, theorem \, for P^1(-|-))$,
3. $P^s(x \mid y) = \dfrac{P^s(y \mid x) \cdot P(x)}{P(y)} \triangleleft P(y) \triangleright P(x)$ $\quad (Bayes' \, theorem \, for P^s(-|-))$,
4. $P^{e/a}(x \mid y) = \dfrac{P^{e/a}(y \mid x) \cdot P(x)}{P(y)} \triangleleft P(y) \triangleright \mathtt{pf}_a(x)$ $(Bayes' \, theorem \, for$ $P^{e/a}(-|-))$.

Proof. Version 1: derive $P^0(x \mid y) = P(x \wedge y)/P(y) \stackrel{3.1.3}{=} (P(y \wedge x)/P(y)) \cdot (P(x)/P(x)) = (P(y \wedge x)/P(x)) \cdot (P(x)/P(y)) = (P^0(y \mid x) \cdot P(x))/P(y)$.

Version 2-4: see http://arxiv.org/abs/1307.5173v4. □

3.3 Independence of Events

A valuated Boolean algebra equipped with a valuation P in some signed meadow $\mathbb{M}$ that satisfies all axioms of $BA + Md + Sign + PF_P$ will be called a $K(\mathbb{M}, P)$-*structure*. Given a $K(\mathbb{M}, P)$-structure, two events x and y are said to be independent relative to that structure if $P(x \wedge y) = P(x) \cdot P(y)$ is valid.

Theorem 3.3.1. *Events x and y are independent if and only if $P^0(x \mid y) = P(x) \cdot P^0(y \mid y)$ and equivalently if and only if $P^0(y \mid x) = P(y) \cdot P^0(x \mid x)$.*

Proof. If x and y are independent, then $P^0(x \mid y) = P(x \wedge y)/P(y) = (P(x) \cdot P(y))/P(y) = P(x) \cdot P^0(y \mid y)$, and similarly one finds $P^0(y \mid x) = P(y) \cdot P^0(x \mid x)$.

Conversely, from $P^0(x \mid y) = P(x) \cdot P^0(y \mid y)$ one finds $P(x \wedge y)/P(y) = P(x) \cdot (P(y)/P(y))$, so multiplying both sides by $P(y)$ yields $P(x \wedge y) \cdot (P(y)/P(y)) = P(x) \cdot (P(y)/P(y)) \cdot P(y)$, which implies $P(x \wedge y) = P(x) \cdot P(y)$ by Theorem 3.1.3. □

4 Logical Aspects of Equations for Probability Functions

In this section we provide a completeness result for $BA + Md + Sign + PF_P$ (Sect. 4.1) and discuss the use of a free Boolean algebra as an event space (Sect. 4.2).

4.1 Completeness of $BA + Md + Sign + PF_P$

In [4] it is shown that $Md + Sign$ constitutes a finite basis for the equational theory of signed cancellation meadows. Stated differently: for each equation $t = r$, if $Md + Sign + PF_P + IL \models t = r$ then also $Md + Sign + PF_P \vdash t = r$, where IL is the inverse law defined in Table 4. This fact is understood as a completeness result because a stronger set of axioms would necessarily exclude some meadows that are expansions of ordered fields. In a preceding version of this paper[6] it was shown that the basis theorem extends to the setting with probability functions: if $BA + Md + Sign + PF_P + IL \models t = r$ then also $BA + Md + Sign + PF_P \vdash t = r$.

For the purposes of this paper we prefer to make use of a different completeness result for the same equational theory that allows us to obtain a more intuitively appealing completeness result for the axiom system $BA+Md+Sign+PF_P$. This second completeness result is given in terms of validity of equations relative to a single signed meadow rather than in an elementary class of structures.

We recall the following result [5, Theorem 3.14], where we write $\mathbb{R}_0$ for the meadow that is the expansion of the field of real numbers $\mathbb{R}$ with total inverse operator and $0^{-1} = 0$, and $(\mathbb{R}_0, \mathbf{s})$ for $\mathbb{R}_0$ expanded with the sign function $\mathbf{s}(\_)$.

Theorem 4.1.1. *For an equation $t = r$ in the signature of signed meadows: $(\mathbb{R}_0, \mathbf{s}) \models t = r$ if and only if $Md + Sign \vdash t = r$.*

One can apply this theorem to obtain a simple proof of Theorem 3.1.3: let $\phi(u, v) = 0_{|u|+|v|} \cdot u$. Then $(\mathbb{R}_0, \mathbf{s}) \models \phi(u, v) = 0$, so by Theorem 4.1.1 one obtains $Md + Sign \vdash \phi(u, v) = 0$. Substituting $P(y \wedge x)$ for u and $P(y \wedge \neg x)$ for v and applying Theorem 3.1.1, one derives

$$\begin{aligned} BA + Md + Sign + PF_P \vdash 0 &= \Big(1 - \frac{|P(y \wedge x)| + |P(y \wedge \neg x)|}{|P(y \wedge x)| + |P(y \wedge \neg x)|}\Big) \cdot P(y \wedge x) \\ &= \Big(1 - \frac{P(y)}{P(y)}\Big) \cdot P(y \wedge x), \end{aligned}$$

from which the required result follows immediately.

The same completeness result as Theorem 4.1.1 works for conditional equations (for a proof see http://arxiv.org/abs/1307.5173v4).

Theorem 4.1.2. *For a conditional equation $t_1 = r_1 \wedge \ldots \wedge t_n = r_n \to t = r$ in the signature of signed meadows: $(\mathbb{R}_0, \mathbf{s}) \models t_1 = r_1 \wedge \ldots \wedge t_n = r_n \to t = r$ if and only if $Md + Sign \vdash t_1 = r_1 \wedge \ldots \wedge t_n = r_n \to t = r$.*

A $K(\mathbb{R}_0, P)$-*structure* is a model of $BA + Md + Sign + PF_P$ that contains the meadow of signed reals, $(\mathbb{R}_0, \mathbf{s})$, as the domain of its values. We will write $K(\mathbb{R}_0, P)$ for the class of $K(\mathbb{R}_0, P)$-structures.

Theorem 4.1.1 can be extended to the setting of $K(\mathbb{R}_0, P)$-structures, thus obtaining a satisfactory completeness result for $BA + Md + Sign + PF_P$ (see http://arxiv.org/abs/1307.5173v4 for a proof that depends on Theorem 4.1.2).

Theorem 4.1.3. *The axiom system $BA + Md + Sign + PF_P$ is sound and complete for the equational theory of $K(\mathbb{R}_0, P)$.*[7]

---

[6] http://arxiv.org/abs/1307.5173v1.

[7] More generally, $BA + Md + Sign + PF_P$ is sound for the class of $K(\mathbb{M}, P)$-structures with $\mathbb{M}$ a signed cancellation meadow.

4.2 Using Free Boolean Algebras as Event Spaces

For the purpose of reformulating some elementary aspects of probability theory and statistics the generality of working with arbitrary Boolean algebras is inessential, at least at this initial stage in the development of an equational calculus of probabilities. For that reason we will now introduce several simplifying assumptions:

- A finite set C of constants for events is provided. Elements of C are called primitive events. We will only consider free Boolean algebras generated by the primitive events.
- With BA_C we will denote the equations for Boolean algebra in a signature which is expanded with the constants in C.
- The class of models of $BA_C + Md + Sign + PF_P$ with a free event space over C, $(\mathbb{R}_0, \mathbf{s})$ as its meadow of values, and a probability function P is denoted $K_C(\mathbb{R}_0, P)$. Different structures in $K_C(\mathbb{R}_0, P)$ only differ in the choice (interpretation) of the probability function P.

These assumptions correspond to what is needed for the specification of examples of probabilistic reasoning.

Theorem 4.2.1. *$Md + Sign + BA_C + PF_P$ is sound and complete for the equations of type V that are true in all structures in $K_C(\mathbb{R}_0, P)$. In other words, for t and r terms of sort V: $Md{+}Sign{+}BA_C{+}PF_P \vdash t = r \iff K_C(\mathbb{R}_0, P) \models t = r$.*

Proof. The proof is merely a reformulation of the proof of Theorem 4.1.3. □

5 Multi-dimensional Probability Functions

In this section we provide axioms for multi-dimensional PFs (Sect. 5.1), and discuss a condition for the existence of a particular universal PF (Sect. 5.2).

5.1 Equational Axioms for a Probability Function Family

Let $D = \{a_1, \ldots, a_d\}$ be a finite, non-empty set. The elements of D are referred to as dimensions. With A_D^f we denote the set of finite non-empty sequences of elements of D in which each dimension occurs at most once, and with $\ell(w)$ we denote the length of $w \in A_D^f$. Note that A_D^f is finite. Elements of A_D^f serve as arities of probability functions on a multi-dimensional event space of dimension $\ell(w)$. If $\ell(w) > 1$, then w is written as a comma-separated sequence, e.g. $\ell(a_1, a_3) = 2$ and we write $(a_1, a_3) \in A_D^f$.

Given an event space E and a name P for a probability function, an arity family for D is a subset W of A_D^f that is closed under permutation and under taking non-empty subsequences. Given an arity family W for D, a function family for W consists of a function $P^w : E^{\ell(w)} \to V$ for each arity $w \in W$. A function family for dimension set D, arity family W and function name P is a *probability function family* (PFF) if it satisfies the axioms of Table 6. Because in an arity repetition of dimensions is disallowed, these axioms reduce to what we had already in the case of a single dimension.

Table 6. *PFF*$_{W,P}$, axioms for a PFF with arity family W and name P, where $a \in D$, $k \in \mathbb{N}$, $\boldsymbol{x} = x_1, \dots, x_k$ and $P(y, \boldsymbol{x}) = P(y)$ if $k = 0$, and $w = (a, u) \in W$ with $\ell(w) = k + 1$

$P^{a,v,b,v'}(y_1, x_1, \dots, x_m, y_2, z_1, \dots, z_n) = P^{b,v,a,v'}(y_2, x_1, \dots, x_m, y_1, z_1, \dots, z_n)$ for all $a, b \in D$ and $(a, v, b, v') \in W$, where v, v' can be empty (thus $m = 0, n = 0$)	(27)
$P^a(\top) = 1$	(28)
$P^{a,v}(\top, x_1, \dots, x_{k+1}) = P^v(x_1, \dots, x_{k+1})$	(29)
$P^w(\bot, \boldsymbol{x}) = 0$	(30)
$P^w(y, \boldsymbol{x}) = \lvert P^w(y, \boldsymbol{x})\rvert$	(31)
$P^w(y \vee z, \boldsymbol{x}) = P^w(y, \boldsymbol{x}) + P^w(z, \boldsymbol{x}) - P^w(y \wedge z, \boldsymbol{x})$	(32)

5.2 Existence of a Universal Probability Function

A subset W of A_D^f may or may not have a maximal element under inclusion. If W has a maximal element $\overline{w}$ and if we have a probability function family $(P^w)_{w \in W}$ for W, then $P^{\overline{w}}$ serves as a universal element for the family of probability functions because all other members of it can be found via successive application of the axioms (27)–(30).

As it turns out some PFFs cannot be extended with a universal PF. In the notation of our specification of probability families we will state a specific result that may serve as a necessary condition for the possibility to extend a PFF with a universal element.

Theorem 5.2.1. *Given a set of dimensions $D = \{a, b, c, d\}$, an arity family W for D that satisfies $W \supset \{(b, c),\ (b, d),\ (a, d),\ (a, c)\}$, and a PFF $(P^w)_{w \in W}$, let t be the following term:*

$$t = P^{b,c}(y, z) + P^{b,d}(y, u) + P^{a,d}(x, u) - P^{a,c}(x, z) - P^b(y) - P^d(u).$$

Then, if W has a maximal element, then $-1 \leq t \leq 0$, that is, the following two inequalities must hold for $G_{W,P} = BA + Md + Sign + PFF_{W,P}$:

$$G_{W,P} \vdash t + 1 = \mathbf{s}(t + 1) \cdot (t + 1) \quad \textit{and} \quad G_{W,P} \vdash -t = \mathbf{s}(-t) \cdot -t.$$

Clearly if a PFF for D contains all of $P^{b,c}, P^{b,d}, P^{a,d}, P^{a,c}$ and fails to meet either one of the mentioned inequalities on t, then a universal PF cannot be found for it.

These facts are known as the BCHS (Bell, Clauser, Horne, Shimony) inequalities. Both were formulated and shown in a set theoretic framework for probability theory in [15] and [10, 11], and a straightforward proof is given in [13, Sect. 9.2],[8] which we repeat here.

[8] From this pair of inequalities one can derive the original Bell inequalities from [3]. The key observation of Bell was that quantum mechanics gives rise to the hypothesis that a 4-dimensional event space exists in which a family of joint probabilities for at most two dimensions can be found that violates the inequalities from the theorem.

*Proof (*of Theorem 5.2.1, taken from [13]*).*

$$
\begin{aligned}
P^{b,c,d}(y,z,u) &= P^{a,b,c,d}(x,y,z,u) + P^{a,b,c,d}(\neg x,y,z,u)\\
&\leq P^{a,c}(x,z) + P^{a,d}(\neg x,u)\\
&= P^{a,c}(x,z) + P^{d}(u) - P^{a,d}(x,u), \qquad (33)\\
P^{b,c,d}(\neg y,z,u) &= P^{a,b,c,d}(x,\neg y,z,u) + P^{a,b,c,d}(\neg x,\neg y,z,u)\\
&\leq P^{a,d}(x,u) + P^{a,c}(\neg x,z)\\
&= P^{a,d}(x,u) + P^{c}(z) - P^{a,c}(x,z), \qquad (34)\\
0 \leq P^{b,c,d}(y,\neg z,\neg u) &= P^{b,c}(y,\neg z) - P^{b,c,d}(y,\neg z,u)\\
&= P^{b}(y) - P^{b,c}(y,z) - P^{b,d}(y,u) + P^{b,c,d}(y,z,u). \qquad (35)
\end{aligned}
$$

Combining (33) and (35) yields

$$0 \leq P^{b}(y) - P^{b,c}(y,z) - P^{b,d}(y,u) + P^{a,c}(x,z) + P^{d}(u) - P^{a,d}(x,u). \qquad (36)$$

By (35) and the equality $-P^{c,d}(z,u) + P^{c,d}(\neg z,\neg u) = 1 - P^{c}(z) - P^{d}(u)$,

$$
\begin{aligned}
&0 \leq P^{b,c,d}(\neg y,\neg z,\neg u) = P^{c,d}(\neg z,\neg u) - P^{b,c,d}(y,\neg z,\neg u)\\
&= 1 - P^{b}(y) - P^{c}(z) - P^{d}(u) + P^{b,c}(y,z) + P^{b,d}(y,u) + P^{b,c,d}(\neg y,z,u). \qquad (37)
\end{aligned}
$$

Then from (34) and (37) we get

$$0 \leq 1 - P^{b}(y) - P^{d}(u) + P^{b,c}(y,z) + P^{b,d}(y,u) + P^{a,d}(x,u) - P^{a,c}(x,z). \qquad (38)$$

Inequalities (36) and (38) prove the theorem. □

6 Concluding Remarks

The incentive for this work came from a talk given by professor Ian Evett on the occasion of the retirement of dr. Huub Hardy as a driving force behind the MSc Forensic Science at the University of Amsterdam.[9] That talk illustrated the headway that the Bayesian approach to reasoning in forensic matters has made in recent years. However, Evett also highlighted the conceptual and political problems that may still lie ahead of its universal adoption in the legal process.

In order to improve the understanding of these issues an elementary logical formalization of reasoning with probabilities might be useful. With that perspective in mind we came to the conclusion that in spite of the abundance of introductory texts to probability theory, the development of an axiomatic approach from first principles may yet cover new ground. The formalization of probabilities in terms of equational logic outlined above is intended to serve as a point of departure from which to develop presentations of probability theory that

[9] This meeting took place at Science Park Amsterdam, Friday June 7, 2013 under the heading "Frontiers of Forensic Science", and was organized by Andrea Haker.

may be be helpful when a formal and logically precise perspective on reasoning with probabilities is aimed at.

We acknowledge many discussions with Andrea Haker (University of Amsterdam) regarding the relevance of logically grounded reasoning methodologies in forensic science. We thank both reviewers for their constructive comments.

References

1. Adams, E.: Probability and the logic of conditionals. In: Hintikka, J., Suppes, P. (eds.) Aspects of Inductive Logic, pp. 265–316. North-Holland (1966)
2. Barber, D.: Bayesian Reasoning and Machine Learning. Cambridge University Press, Cambridge (2012). http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.Online. ISBN 0521518148, 9780521518147. On-line version 18 November 2015
3. Bell, J.S.: On the Einstein Podolsky Rosen paradox. Physics **1**(3), 195–200 (1964)
4. Bergstra, J.A., Bethke, I., Ponse, A.: Cancellation meadows: a generic basis theorem and some applications. Comput. J. **56**(1), 3–14 (2013). https://doi.org/10.1093/comjnl/bxs028. https://arxiv.org/abs/0803.3969v3
5. Bergstra, J.A., Bethke, I., Ponse, A.: Equations for formally real meadows. J. Appl. Log. **13**(2, Part B), 1–23 (2015). https://doi.org/10.1016/j.jal.2015.01.004. https://arxiv.org/abs/1310.5011v4
6. Bergstra, J.A., Middelburg, C.A.: Inversive meadows and divisive meadows. J. Appl. Log. **9**(3), 203–220 (2011). https://arxiv.org/abs/0907.0540v3
7. Bergstra, J.A., Middelburg, C.A.: Division by zero in involutive meadows. J. Appl. Log. **13**(1), 1–12 (2015). https://arxiv.org/abs/1406.2092
8. Bergstra, J.A., Ponse, A.: Fracpairs and fractions over a reduced commutative ring. Indag. Math. **27**, 727–748 (2016). https://doi.org/10.1016/j.indag.2016.01.007. https://arxiv.org/abs/1411.4410v2
9. Carlström, J.: Wheels - on division by zero. Math. Struct. Comput. Sci. **14**(1), 143–184 (2004). https://doi.org/10.1017/S0960129503004110
10. Fine, A.: Joint distributions, quantum correlations, and commuting observables. J. Math. Phys. **23**, 1306–1310 (1982). https://doi.org/10.1063/1.525514
11. Fine, A.: Hidden variables, joint probability, and the Bell inequalities. Phys. Rev. Lett. **48**(5), 291–295 (1982). https://doi.org/10.1103/PhysRevLett.48.291
12. Milne, P.: Bruno de Finetti and the logic of conditional events. Br. J. Philos. Sci. **48**, 195–232 (1997)
13. de Muynck, W.M.: Foundations of quantum mechanics, an empiricist approach. In: Fundamental Theories of Physics, vol. 127. Kluwer Academic Publishers (2002). Pre-publication version: http://www.phys.tue.nl/ktn/Wim/Kluwerbookprepversion.pdf
14. Padmanabhan, H.: A self-dual equational basis for Boolean algebras. Canad. Math. Bull. **26**(1), 9–12 (1983)
15. Rastall, P.: The Bell inequalities. Found. Phys. **13**(6), 555–570 (1983)
16. dos Reis, T.S., Gomide, W., Anderson, J.A.D.W.: Construction of the transreal numbers and algebraic transfields. IAENG Int. J. Appl. Math. **46**(1), 11–23 (2016). http://www.iaeng.org/IJAM/issues_v46/issue_1/IJAM_46_1_03.pdf