Conditional Values in Signed Meadow Based Axiomatic Probability Calculus

Bergstra, J.A.

Publication date
2016

Document Version
Final published version

Conditional Values in Signed Meadow Based Axiomatic Probability Calculus

Jan A. Bergstra

Informatics Institute, University of Amsterdam*

November 4, 2016

Abstract

An equational axiomatisation of probability functions for one-dimensional event spaces in the language of signed meadows is expanded with conditional values and configurations. Assuming the presence of a probability function, equational axioms are provided for expectation value, variance, covariance, and correlation squared, each for conditional values, and for expected utility of configurations. Finite support summation is introduced as a binding operator on meadows which simplifies formulating requirements on probability mass functions with finite support. Conditional values are related to probability mass functions and to random variables. The definitions are reconsidered in a higher dimensional setting.

Keywords and phrases: Boolean algebra, signed meadow, probability function, probability mass function, conditional value.

1 Introduction

In [5] a proposal is made for a loose algebraic specification probability functions in the context of signed meadows. This specification is referred to as BA + Md + Sign + PF_P thereby containing abbreviations for "Boolean algebra", "meadows", "sign function", and "a single probability function" with name P. The specification BA + Md + Sign + PF_P will be recalled in detail below. Probability function will be abbreviated to PF.

The objective of this paper is to proceed on the basis of the results of [5] and to provide an account of some basic elements of probability calculus including expectation value, probability mass function, variance, covariance, correlation, independence, configuratio, sample space, and random variable. I refer to this particular approach to presenting probability calculus as signed meadow based axiomatic probability calculus. The presentation makes use of the special properties of meadows, most notably $1/0 = 0$, without hesitation and the presentation is organised so as to fit best with the equational and axiomatic setting of meadows.

*Email: `j.a.bergstra@uva.nl`, `janaldertb@gmail.com`

A conventional ordering of the introduction of concepts in probability theory is as follows: (i) given a sample space S, an event space E is introduced as a subset of the power set of S. (ii) Then probability functions are defined over event spaces and (iii) discrete random variables are introduced as real functions on S with a countable support. Given these ingredients, (iv) expectation value and variance are defined for discrete random variables and covariance, and correlation are defined for pairs of random variables. (v) Subsequently probability mass functions are derived from random variables, (vi) multivariate discrete random variables are introduced as vectors of random variables on the same event space, and (vii) joint probability mass functions are derived from multivariate random variables, (viii) independence is defined from joint probability mass functions, and (ix) marginalisation is defined as a transformation on joint probability mass functions. (x) The development of concepts and definitions is redone for the continuous case with probability distributions replacing probability mass functions and (continuous) random variables replacing discrete random variables.

Below the same aspects will be discussed under quite restrictive conditions and in a different order. The technical core notion for this paper is a probability mass function with finite support. This is a real function taking nonzero values for finitely many arguments, that is the elements of its support, and such that the sum of the nonzero values adds up to one. All other notions that are discussed below will be linked with finitely supported probability mass functions. As a conceptual cornerstone the space (sort) of conditional values is introduced.

The roles of sample space and random variables are given a less central position than in a conventional exposition. Both topics are dealt with in the Appendix, while conditional values will play the role that is conventionally played by random variables.

As a consequence of these choices the definition for the expectation value, and the definitions of (co)variance and correlation which directly depend on expectation values, will be repeated in three different settings: (i) for probability mass functions with finite support, (ii) for an event space equipped with a sort of conditional values in addition to a probability function, and (iii) for a multidimensional event space equipped with a sort of conditional values in addition to a family of multivariate probability functions.

By choosing this order of presentation an adequate match is obtained with meadow based equational axiomatisations. By having an account of probability mass functions which merely treats those as a particular class of real functions that part of the story is made independent of probability theory and the impact and virtue of working with meadows can be assessed in isolation. By defining expectation values and derived quantities on conditional values over an event structure the incentive for introducing a sample space is avoided, thus avoiding an incentive to introduce a subsort of samples for the sort of events, and thereby maintaining the simplicity of the use of a loose equational specification for probability functions.[1]

By making a clear distinction between the single dimensional case and the multi-dimensional case, the potential nonexistence of joint probability functions, and how that potential nonex-

[1]This argument also motivates the introduction of a sort CV for conditional values. Avoiding a new sort, and having CVs as an extension of the sort V of values instead, seems profitable at first sight. But then the sign of a conditional value must be defined, if sign is to remain a total function, and there seems to be no natural definition available. In the presence of a probability function, the sign of a conditional value may be identified with the sign of its expectation value. Doing so is unnatural, however, why would sign depend on the choice of a probability function, and technically complicated, as new equations for sign will be needed, the current one's being incompatible with the presence of conditional values.

istence relates to equational axiomatisation, is fully taken into account.

1.1 Survey of the paper

As a first step the notion of a probability mass function (PMF) with finite support is introduced and its formal specification in the setting of meadows is provided with the help of a new operator called finite support summation (FSS). By default a PMF is assumed to be univariate. A PMF with finite support is a function from reals to reals which has finitely many non-zero values at most and for which the sum of its non-zero values adds up to 1.[2]

In addition to FSS multivariate versions of FFS are defined for all arities, and marginalisation is defined as a family of transformations from an FSS with more than one argument to an FSS with a smaller number of arguments. Expectation value and variance are defined as functionals on (univariate) PMFs and covariance and correlation are defined as functionals on bivariate PMFs.

Having developed an account of PMFs independently of axioms for probability functions, Section 4 proceeds with a recall from [5] of the combination of an event space (a Boolean algebra) and value space (a meadow), and the equational specification of a probability function (PF). Two versions of Bayes' rule are considered an the relative position of these statements w.r.t. the axioms is determined.

Independently of PFs a conditional operator is applied to events and the results of the operator are collected in a new sort CV of so-called conditional values, which is comes with the structure of a meadow. Assuming that the Boolean algebra serving as an event space has more than two distinct elements the sort CV constitutes a non-cancellation meadow.

Thinking in terms of outcomes of a probabilistic process one may assume that the process produces as an outcome an entity of some sort. Events from an event space $\mathbb{E}$ represent assessments about the outcome. It is plausible that besides Boolean assessments also values from a meadow (for instance reals) are considered attributes of an outcome. A CV directly relates values from value space to events from event space. In the presence of a PF, expectation value is specified for CVs, by means two equational axioms.

According to [5] the equations of BA + Md + Sign + PF_P constitute a finite equational basis for the class of Boolean algebra based, real valued PFs and the proof theoretic results, viz. soundness and completeness, concerning signed meadows of [3, 4] extend to the case with Boolean algebra based PFs. The axiom system BA + Md + Sign + PF_P may be viewed as a merely a particular formalisation of Kolmogorov's axioms for probability theory phased in the context of (involutive) meadows and the completeness result asserts the completeness of this particular formalisation w.r.t. its (second order) semantics, that is its standard model. Below this axiomatisation is extended with CVs and E_P and a corresponding completeness result holds.

Configurations are introduced as parallel configurations of objects subject to conditions made up from events. This provides a configuration space. Configurations are equipped with CVs that play the role of values which can be uniformly extracted from a configuration by means of the utility function. It is plausible to design a CS tailored to a specific context. Two

[2] A PMF with finite support may be understood as a probability distribution over a finite set of reals.

examples of such designs are provided.

Multivariate CVs are vectors of CVs. From [5] the specification for multi-dimensional PFs relative to an arity family is imported. Multi-dimensional PFs can serve as joint probability distributions obtained from a multivariate CV and a multidimensional PR over an arity family, provided the arity family is sufficiently rich to contain the type of the multivariate RV at hand. The multidimensional case is important because it admits a precise formalisation of cases where the existence of joint probability functions cannot be taken for granted. nition of covariance and correlation can be found by first extracting a PMF from the respective RVs.

1.2 Applying equational logic

Making use of first order logic as a tool for the exposition of a particular subject creates certain difficulties which merit some reflection in advance. First of all, by working in first order equational logic, I intend to provide and support a new axiomatic approach to the elementary theory of probability. The objective of formalisation and axiomatic style below is not to have all work done with the overarching intention to avoid mistakes. Rather the idea is to use the axiomatic approach to obtain maximal clarity about assumptions, working hypotheses, patterns of reasoning, and patterns of calculation. My impression is that by applying first order equational logic to the presentation of probability calculus additional insight is gained.

A complication, however, which easily pops up when applying formal methods, is a phenomenon which I will refer to as *overformalisation*. In order to develop a valid presentation from the point of view of formal logic one tends to write all assertions in formal notation and to have only rudimentary basics explained in conventional mathematical terms. Now it may easily come about that in the presence of formalised fragments of text neighbouring fragments written in conventional mathematical style are viewed as deficient. Overformalisation is likely to occur in a text if, for the sake of consistency of presentation while detrimental to its readability, each part of the text is cast in a formal shape even when no additional clarity can be gained by doing so. An example of this phenomenon may occur if different notations are used for a constant and its meaning. In order to avoid this instance of overformalisation I will not distinguish between constants and functions concerning meadows and their mathematical counterparts. For instance the convention to write $\underline{0}$ for the constant symbol that is interpreted by the value 0 is not adopted and the reader is supposed to infer from the context whether an occurrence of 0 refers to a formal constant symbol or to a specific element of some mathematical domain. Another instance of overly formal presentation occurs if $\mathbb{R}_0 \models t = r$ is written in cases where ordinary mathematics suggest writing $t = r$. If no confusion is expected simply writing $t = r$ is preferred. On the other hand the sort E of events will be distinguished from the corresponding domain $||\mathbb{E}||$ in a particular event structure $\mathbb{E}$ and the sort of values V for meadows will be distinguished from the domain $||\mathbb{M}||$ of some meadow $\mathbb{M}$, and a specific probability function will be referred to as $\mathbb{P}$ while in formalised notation its name is P.

Below equational logic will applied with the following objectives in mind: (i) to demonstrate that an axiomatic approach in terms of equational logic to elementary probability calculus is both feasible and attractive, (ii) to illustrate the compatibility of an axiomatic approach to probability calculus with conventional mathematical style and notation, and (iii) to provide optimal clarity about the assumptions which underly the definition of key concepts, while (iv)

$$(x + y) + z = x + (y + z) \quad (1)$$
$$x + y = y + x \quad (2)$$
$$x + 0 = x \quad (3)$$
$$x + (-x) = 0 \quad (4)$$
$$(x \cdot y) \cdot z = x \cdot (y \cdot z) \quad (5)$$
$$x \cdot y = y \cdot x \quad (6)$$
$$1 \cdot x = x \quad (7)$$
$$x \cdot (y + z) = x \cdot y + x \cdot z \quad (8)$$
$$(x^{-1})^{-1} = x \quad (9)$$
$$x \cdot (x \cdot x^{-1}) = x \quad (10)$$
$$1_x = x \cdot x^{-1}$$
$$0_x = 1 - x \cdot x^{-1}$$
$$x^2 = x \cdot x$$
$$x/y = x \cdot y^{-1},$$
$$x \lhd y \rhd z = 1_y \cdot x + 0_y \cdot z.$$
$$\frac{x}{y} = x/y,$$

Table 1: Md: axioms for meadows and defining equations for abbreviations

using meadows as a tool throughout the presentation.

2 Meadows and finite support summation

Numbers will be viewed as elements of a meadow rather than as elements of a field. For the introduction of meadows and elementary theory about meadows I refer to [8, 3, 4] and the papers cited there. I will copy the tables of equational axioms for meadows and for the sign function which plays a central role below. With $(\mathbb{R}_0, \mathbf{s})$ the expansion of the meadow $\mathbb{R}_0$ with the sign function is denoted. The following completeness result was obtained in [4].

Theorem 1. *A conditional equation in the signature of signed meadows is valid in* $(\mathbb{R}_0, \mathbf{s})$ *if and only if it is provable from the axiom system Ms+Sign.*

The set of axioms in Table 1 specifies the class of meadows and the axioms in Table 2 specify the sign function. Following [3], a meadow that satisfies the (nonequational) implication IL from Table 3 is called a cancellation meadow. Table 2 also includes definitions of the ordering relations in terms of the sign function. Ordering comes in two versions: as a relation and as a function to values with 1 representing truth and 0 representing falsehood. Together these tables contain the axioms for signed meadows, as well as a catalogue of well-known operations on

$$\mathbf{s}(1_x) = 1_x \tag{11}$$
$$\mathbf{s}(0_x) = 0_x \tag{12}$$
$$\mathbf{s}(-1) = -1 \tag{13}$$
$$\mathbf{s}(x^{-1}) = \mathbf{s}(x) \tag{14}$$
$$\mathbf{s}(x \cdot y) = \mathbf{s}(x) \cdot \mathbf{s}(y) \tag{15}$$
$$0_{\mathbf{s}(x)-\mathbf{s}(y)} \cdot (\mathbf{s}(x+y) - \mathbf{s}(x)) = 0 \tag{16}$$
$$|x| = \mathbf{s}(x) \cdot x$$
$$x < y \equiv_{\mathtt{def}} \mathbf{s}(y - x) = 1$$
$$x \leq y \equiv_{\mathtt{def}} \mathbf{s}(\mathbf{s}(y - x) + 1) = 1$$

Table 2: Sign: axioms for the sign operator

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

Table 3: IL: inverse law

these, which are introduced as named derived (that is having an explicit definition) operators.

The following useful characterisation of x being non-negative is valid in $(\mathbb{R}_0, \mathbf{s}) : 0 \leq x \iff x = \mathbf{s}(x) \cdot x$. And therefore its provability follows from [4]:

Proposition 1. $Md + Sign \vdash 0 \leq x \iff x = \mathbf{s}(x) \cdot x \iff 0 \leq_v x = 1.$

2.1 Finite support summation (FSS)

Given a meadow $\mathbb{M}$ and a term t in which variable x may or may not occur it may be useful to determine the summation of all substitutions (or rather interpretations) $[v/x]t$ with $v \in ||\mathbb{M}||$. This sum is unambiguously defined, however, if the support in $\mathbb{M}$ of $\lambda x.t$ is finite, that is if there are only finitely many values $v \in ||\mathbb{M}||$ such that $[v/x]t$ is nonzero.

The expression $\sum_x^\star t$ denotes in M the sum of all $[v/x]t$ if at most finitely many of these substitutions $[v/x]t$ yield a non-zero value and 0, otherwise.

The $\sum_x^\star$ operator will be referred to as finite support summation (FSS). At this stage we have little information about the logical properties of this binding operation on terms but it is semantically unproblematic, being well-defined in each meadow, and it will be used below for presenting some rigorous and yet intuitively appealing definitions.

We first notice some technical facts concerning FSS, assuming the interpretation of equations is performed in an arbitrary cancellation meadow M.

1. $\lambda x.t$ has finite support iff $\lambda x.t/t$ has finite support.
2. $\sum_x^\star 1_t = |\sum_x^\star 1_t|$.

3. $\sum_x^\star 0 = 0$.
4. $\sum_x^\star 1 = 0$. To see this first notice that in an infinite meadow, including the meadows of characteristic 0, 1 is nonzero for infinitely many x and thus $\sum_x^\star 1 = 0$. In a finite meadow of nonzero characteristic (say p) $\sum_x^\star 1$ counts the cardinality of the structure, which must be a multiple of p and therefore vanishes.
5. $\sum_x^\star (1_x) = 0$ if and only if $\mathbb{M}$ is infinite.
6. $\sum_x^\star (1_x) = -1$ if and only if $\mathbb{M}$ is finite.
7. $\sum_x^\star (0_x) = 1$.
8. $\sum_x^\star (t + 0_x) - \sum_x^\star t = 1$ if and only if $\lambda x.t$ has finite support.
9. $\sum_x^\star (t + 0_x) = \sum_x^\star t$ if and only if $\lambda x.t$ has infinite support.
10. Let the context $s[\,]$ be given by: $s[t] = (1_{(\sum_x^\star (1_t))} \lhd (\sum_x^\star (t+0_x) - \sum_x^\star t) \rhd 1$, then $s[t] = 1$ if and only if the support of $\lambda x.t$ is nonempty, and otherwise $s_t = 0$.
11. with $s[\,]$ as in item 10: $s[0_{s[t]}] = 1$ if $\lambda x.t$ takes value zero somewhere, and $s[0_{s[t]}] = 0$ otherwise.
12. If $x \notin FV(t)$ then $\sum_x^\star (r \cdot t) = (\sum_x^\star r) \cdot t)$.
13. If $x \notin FV(t)$ then $\sum_x^\star (x \cdot 0_{t-x}) = t$.
14. If both $\lambda x.t$ and $\lambda x.t$ have finite support then $(\sum_x^\star t) + (\sum_x^\star r) = \sum_x^\star (t + r)$.

2.2 Multivariate finite support summation

The multivariate case of FSS operations requires separate definitions for each number of variables because a stepwise reduction to the definition for the univariate case is unfeasible. To demonstrate this difficulty we consider the bivariate case only, the case with three or more variables following the same pattern. In a meadow $\mathbb{M}$, $\sum_{x,y}^\star t$ produces 0 if for infinitely may pairs of values $a, b \in ||\mathbb{M}||$ the value of $[a/x][b/y]t$ is nonzero, otherwise it produces the sum of the finitely many nonzero values thus obtained.

The need for expressions of the form $\sum_{x,y}^\star t$ transpires from an elementary example, which demonstrates that a 2-dimensional FSS cannot be simply reduced to a composition of 2 occurrences of a 1-dimensional FFS. Let $t(x, y) = 0_x \cdot 0_y + 0_{1-x}$. Because $t(1, y) = 1$ for all y, $t(x, y)$ is nonzero on infinitely many pairs of values, so that

$$\sum_{x,y}^\star t(x, y) = 0.$$

Now notice that $\sum_y^\star t(0, y) = 1$, $\sum_y^\star t(1, y) = 0$, and if $x \neq 0 \wedge x \neq 1$, $\sum_y^\star t(x, y) = 0$. It follows that

$$\sum_x^\star \sum_y^\star t(x, y) = 1.$$

2.3 Countable support summation

Another summation operator that may be of relevance for probability calculus is countable support summation (CSS), written $\sum_x^{\star\star} t$, which delivers the sum of the absolute values of the values that $\lambda x.t$ takes on its support provided the support of $\lambda x.t$ is countable and the sum converges, in all other cases $\lambda x.t = 0$.

The definition of CSS works correctly for each expression t over a signed meadow provided (i) it is well-defined whether or not a countable sum of absolute values has a limit, and (ii) limits of countable sums of absolute values are unique. A sufficient criterion for these constraints is that a signed meadow is Archimedian. However, in view of the fact that the content of the paper depends on the application of FSS only when adopting Definition 2 below as a starting point rather than Definition 1, no attention is paid below to the details of properly defining CSS in this paper.

2.4 Finite support multiplication

A generalised multiplication operator $\prod_x^{\star} t$, which is complementary to FSS, can be defined by assuming value 0 if $\lambda x.(t-1)$ has finite support and by denoting the product of all values of $\lambda x.t$ that differ from 1 otherwise. This operator is named FSM (finite support multiplication). FSM satisfies $t = 0 \to \prod_x^{\star} t = 0$. FSM will not play a role in this paper.

Many questions can be stated about FSS, CSS and FSM which have not yet been investigated. Working over $\mathbb{Q}_0$ or $\mathbb{R}_0$ for instance one may ask about the decidability of (fragments of) the equational logic obtained after extending the language with FSS.

2.5 Representing functions by expressions

There are several ways in which an expresion can be used to denote a function of type $V \to V$. I will now briefly discuss four such options.

A common approach is to extend the expression language with lambda abstraction and to admit $\lambda x.t$ as an expression denoting the function which maps $v \in V$ to $[v/x]t$, i.e. the result of substituting v for x in t. Unfortunately lambda-abstraction seems to be still unfamiliar to an audience outside logic and computing.

Another option is to consider a pair $(x;t)$ to represent the same function. The difference with lambda abstraction is that alpha conversion, i.e. renaming the bound variable x to another variable name which does not occur in t produces a different representation.

A third option is always to insist that the variable for which an argument is to be substituted is explicitly mentioned in a binding operator and to leave the pair $(x;t)$ implicit in a description.

Finally it is an option to fix in advance one or more variables, say $x_1, x_2, \ldots$ and to assume that if t is to represent a function with n arguments, a fact which must be inferred from the context at hand, it represents the function $(x_1, \ldots, x_n; t)$, i.e. that x_i serves as its i-th argument.

When FSS is to be applied to a function these four options lead to different notations:

$\sum(\lambda x.t)$, $\sum(x;t)$, $\sum_x t$, and $\sum_{x_1} t$, respectively. There is no need to choose a single convention from these four options and below it is supposed to be clear from the context which one of these conventions is used in each particular case.

Except when lambda abstraction is used the notion of a key variable of a function representation applies. The key variable of $(x;t)$ is x, also if $(x;t)$ appears only implicitly, and in the fourth case x_i serves as the i-th key variable.

When denoting expressions t and $t(x)$ have the same meaning except that in a context where $t(x)$ is used the convention is implicitly introduced that $t(r)$ is supposed to be a shorthand for $[r/x]t$. Writing $t(x)$, however, is not supposed to imply the existence of a free occurrence of x in t.

3 PMFs with finite support

The main application of FSS in this paper is to enable the following definition of what it means for a term to represent a finitely supported probability mass function. Probability mass function will be abbreviated as PMF.

Definition 1. *A term t coupled with a key variable x represents (in accordance with the third option from the listing in Paragraph 2.5 is used) a PMF, written as $\lambda x.t$, in a signed meadow $\mathbb{M}$ if and only if these two conditions are met:*

1. $\mathbb{M} \models \sum_x^{\star\star} t = 1$, *and*
2. $\mathbb{M} \models t = |t|$.

Finitely supported PMFs constitute a special case of PMFs:

Definition 2. *A term t coupled with a key variable x represents a PMF with finite support (PMF with finite support, finitely supported PMF, written as $\lambda x.t$) in a signed meadow $\mathbb{M}$ if and only if these two conditions are met:*

1. $\mathbb{M} \models \sum_x^{\star} t = 1$, *and*
2. $\mathbb{M} \models t = |t|$.

Although this definition of a term representing a *PMF with finite support* is technically unproblematic the use of terminology from probability theory requires some justification. Indeed PMFs occur in probability theory where these comprise precisely all nonnegative functions from reals to reals with a countable support so that the sum of all non-zero values equals 1. With this fact in mind, and working in the signed meadow $\mathbb{R}_0(\mathbf{s})$, the two requirements of Definition 2 indeed guarantee that the function represented by $(t;x)$ is a PMF with finite support according to standard terminology.

The property of being a representative of a finitely supported PMF is quite sensitive to the meadow at hand. This fact is best illustrated with an example. Consider the expression

$$t(x) = \frac{1}{4} \cdot 0_{x^2-2} \cdot ((1 + \mathbf{s}(x)) \cdot x + ((1 - \mathbf{s}(x)) \cdot (2 - x))).$$

In $\mathbb{R}_0$ the function description $(x;t)$ represents a finitary PMF. To see this notice that it takes non-zero values only in $-\sqrt{2}$ and $\sqrt{2}$ where it has values $1-1/2\sqrt{2}$ and $1/2\sqrt{2}$ respectively while in $\mathbb{Q}_0$ it is not the case that $(t;x)$ represents a finitely supported PMF because $t(x)$ vanishes for all x with the implication that $\sum_x^\star t = 0$. On the other hand when considering $t'(x) = t(x) + 0_x$ one finds that $(x;t')$ represents a finitely supported PMF in $\mathbb{Q}_0$ while it fails to do so in $\mathbb{R}_0$.

3.1 Multivariate PMFs with finite support

A joint PMF with finite support of arity n is a function, say denoted with $(x_1,\ldots,x_n;F(x_1,\ldots,x_n))$ from $\mathbb{R}_0^n$ to $\mathbb{R}_0$ which satisfies these two conditions:

1. $\sum^\star_{x_1,\ldots,x_n} F(x_1,\ldots,x_n) = 1$, and
2. for all $x_1,\ldots,x_n \in \mathbb{R}_0^n$, $F(x_1,\ldots,x_n) = \mathbf{s}(F(x_1,\ldots,x_n)\cdot F(x_1,\ldots,x_n)$.

3.2 Explicit definition of joint PMFs with finite support

I will consider the bivariate case as an example. Assuming that information about the graph of a joint PMF with finite support, with exception of argument vectors for which the result vanishes, is encoded in a set of triples: $\{(y_{1,1},y_{2,1},z_1),\ldots,(y_{1,n},y_{2,n},z_n)\}$, a corresponding function expression F for the same joint PMF with key variables x_1 and x_2 is as follows:

$$F(x_1,x_2) = \sum_{i=1}^{n} (0_{(x_1-y_{1,i})^2+(x_2-y_{2,i})^2} \cdot z_i).$$

3.3 Marginalisation and independence

Given a finitely supported joint PMF G with n variables $x_1,\ldots,x_n$, marginalisation can be defined to each subset $x_{i_1},\ldots,x_{i_k}$ with $1 \leq i_1 < \ldots < i_k \leq n$. Let $x_{j_1},\ldots,x_{j_{n-k}}$ be an enumeration without repetition of the variables in $x_1,\ldots,x_n$ that are not listed in $x_{i_1},\ldots,x_{i_k}$, then $G_{(i_1,\ldots,i_k)}$ represents a joint PMF with k variables $x_{i_1},\ldots,x_{i_k}$ as follows:

$$G_{(i_1,\ldots,i_k)}(x_{i_1},\ldots,x_{i_k}) = \sum^{\star}_{x_{j_1}\ldots,x_{j_{n-k}}} G(x_1,\ldots,x_n).$$

For a bivariate PMF $G(x,y)$ independence is defined as independence of its two marginalisations.

$$IND(G) \equiv_{def} \forall x,y \in V.G(x,y) = G_{(1)}(x)\cdot G_{(2)}(y).$$

3.4 Expectation, (co)variance, and correlation for PMFs

Now $F(x)$ is assumed to be a term representing a finite support PMF with x as the key variable, while $G(x,y)$ represents a joint PMF with finite support with x as the first and y

as the second key variable. Two PMFs $G_{(1)}$ and $G_{(2)}$ are derived from G by marginalization: $G_{(1)}(x) = \sum_y^\star G(x,y)$ and $G_{(2)}(y) = \sum_x^\star G(x,y)$.[3]

Expectation value. The expectation value $E_{pmf}(F)$ of F is given by:

$$E_{pmf}(F) = \sum_x^\star (x \cdot F(x)).$$

Variance. The variance of F is defined by:

$$VAR_{pmf}(F) = \sum_x^\star (x^2 \cdot F(x)) - E_{pmf}(F)^2.$$

Covariance. For G the covariance is given by:

$$COV_{pmf}(G) = \sum_{x,y}^\star (x \cdot y \cdot G(x,y)) - E_{pmf}(G_{(1)}) \cdot E_{pmf}(G_{(2)}).$$

Correlation. For G the square of correlation is given by:

$$CORR^2_{pmf}(G) = \frac{COV_{pmf}(G)^2}{VAR_{pmf}(G_{(1)}) \cdot VAR_{pmf}(G_{(2)})}.$$

The explicit definition of the square of correlation is given in order not to burden the present exposition paper with the equational specification of a square root operator, which is an option that constitutes an entirely separate topic.[4] These definitions admit a justification in hindsight on the basis of the conventional use of the defined notions. Details of that justification have been referred to a footnote.[5]

4 Event spaces and probability functions

From [5] I will recall equations for Boolean algebras, (signed) meadows, and probability functions.[6] A Boolean algebra $(B, +, -, \_, 1, 0)$ may be defined as a system with at least two

[3] $E_{pmf}(-)$ and the three other operators for variance, covariance, and correlation squared, as defined below are equipped with subscripts in order to avoid the confusion that may arise if on thinks in terms of the argument F being a random variable which it is not.

[4] For the definition of correlation via $CORR_{pmf}(G) = \sqrt{(}CORR^2_{pmf}(G))$ the availability is needed of a totalised square root function. Square root can be made total by writing $\sqrt{(}-x) = -\sqrt{(}x)$. Details of this version of the square root function including appropriate equational axioms can be found in [3]. Without proof we mention that the completeness result of Proposition 1 carries over to the setting of square root meadows. The proof combines methods and results of [3, 4].

[5] Let $(t; x)$ represent a non-negative real function $\lambda x.t$ with finite support, say S. S may be viewed as a sample space and then id_S, the identity function of type $S \to \mathbb{R}_0$ qualifies as a random variable, say X, in conventional terminology. The finite powerset of S serves as an event space $\mathbb{E}$. Now choose a probability function P generated by $P(\{s\}) = F(s)$ for $s \in S$. It follows that $P(X = x) = F(x)$, X being a restriction of identity to the reals. Now $E_{pmf}(F) = E_P(X) = \sum_{s \in S}(X(s) \cdot P(X = s)) = \sum_x^\star (x \cdot F(x))$, and $VAR_{pmf}(F) = E_P(X^2) - (E_P(X))^2 = \sum_{s \in S}(x^2 \cdot P(X = s)) - (E(X))^2 = \sum_x^\star (x^2 \cdot F(x)) - (E(X))^2$.

[6] In [7] probabilistic choice is formalised with the meadow of reals as a number system. Though entirely different from the current paper, the equations in that paper demonstrate, just as well as the equations in

$$(x \vee y) \wedge y = y \tag{17}$$
$$(x \wedge y) \vee y = y \tag{18}$$
$$x \wedge (y \vee z) = (y \wedge x) \vee (z \wedge x) \tag{19}$$
$$x \vee (y \wedge z) = (y \vee x) \wedge (z \vee x) \tag{20}$$
$$x \wedge \neg x = \bot \tag{21}$$
$$x \vee \neg x = \top \tag{22}$$

Table 4: BA: a self-dual equational basis for Boolean algebras

elements such that $\forall x, y, z \in B$ the well-know postulates of Boolean algebra are valid. In order to avoid overlap with the operations of a meadow, Boolean algebras are equipped with notation from propositional logic, thus consider $(B, \vee, \wedge, \neg, \top, \bot)$ and adopt the axioms as presented in Table 4. In [15] it was shown that the axioms in Table 4 constitute an equational basis for the equational theory of Boolean algebras.

In the setting of probability functions the elements of the underlying Boolean algebra are referred to as events.[7] I will use "value" to refer to an element of a meadow,[8] and a probability function is a particular kind of valuation from events to the values in a signed meadow.[9]

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 only a single name for a probability function, the function symbol P.[10] Table 5 provides axioms for a probability function as well as a defining equation for conditional probability as a derived operator. Table 5 makes use of inversive notation.[11] Together with the axioms for signed meadows and for Boolean algebras we find the following set of axioms: BA+Md+Sign+PF$_P$.

Table 6 provides explicit definitions of some useful conditional probability operators made total by choosing a value in case the condition has probability 0.

Table 5 below, a reasonable compatibility between the requirements of probability calculus and the treatment of division in a meadow.

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

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

[9] The terminology of probability theory has variations, for instance in [9] a probability function is referred to as a probability law while constraints on probability laws are referred to as axioms.

[10] 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.

[11] The term inversive notation was coined in [6]. It stands in contrast with divisive notation that makes use of a two place division operator symbol which is provided as a derived operation and which itself conventionally appears in two forms.

$$P(\top) = 1 \quad (23)$$

$$P(\bot) = 0 \quad (24)$$

$$P(x) = |P(x)| \quad (25)$$

$$P(x \vee y) = P(x) + P(y) - P(x \wedge y) \quad (26)$$

Table 5: PF_P: axioms for a probability function with name P

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

$$P^1(x \mid y) = P(x \wedge y) \cdot P(y)^{-1} \lhd P(y) \rhd 1$$

$$P^s(x \mid y) = P(x \wedge y) \cdot P(y)^{-1} \lhd P(y) \rhd P(x)$$

Table 6: Conditional probability operators

4.1 Soundness and completeness of axioms for probability functions

The reader is assumed to be familiar with the concept of a PF, say $\mathbb{P}$ with name P, on an event space $\mathbb{E}$, where $\mathbb{P}$ is supposed to comply with the informal Kolmogorov axioms of probability theory. In the light of assuming the presence of reals, sets, and measures on sets as given, the Kolmogorov axioms are more easily understood as providing a mathematical definition, that is a set of requirements, governing which functions are considered PFs than as constituting a formal system of axioms.[12]

Just as in the cases of number theory or set theory a seemingly unambiguous mathematical idea may give rise to a plurality of formalisations. The axiom system BA+Md+Sign+PF_P serves as a formalisation of the Kolmogorov's axioms for probability functions. Different formalisations of the Kolmogorov are specified in Section 6 where a family of n-dimensional formalisations, for $n \in \mathbb{N}$, of PFs is displayed each of which formally captures the same ideas on probability though in a somewhat different manner.

A probability function structure (PF structure) over an event space E is a two sorted structure having E (events) and V (values) as sorts with E interpreted by a Boolean algebra and V interpreted as the real numbers, enriched with a probability function $\mathbb{P}$ from E to V. The Kolmogorov axioms specify precisely which functions are probability functions. I will assume that V is the domain of the meadow of reals, i.e. that the meadow version of real numbers is used. With $EPV(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), P)$ the class of PF structures over a fixed event structure $\mathbb{E}$ is denoted, with values taken in $||\mathbb{R}_0(\mathbf{s})||$. For a specific PMF $\mathbb{P}$ the pertinent

[12]In [17] an extensive survey is provided of the history leading up to Kolmogorov's choice of axioms, and to his claim that these axioms are what probability is about. In fact Table 5 does not take the 6th axiom into account which asserts that if $(e_i)_{i\in\mathbb{N}}$ is an infinite descending chain of events such that only $\bot$ is below each element of the chain, then $\lim_{i\to\infty} P(e_i) = 0$. A closer resemblance with Kolmogorov's original axioms is found if equation 24 is omitted and equation 26 is replaced by the conditional equation $e \wedge f = \bot \to P(e \vee f) = P(e) + P(v)$. These modifications produce a logically equivalent axiom system. The equations of Table 5 are preferred because of the equational form.

$$P(\top) = 1$$
$$P(\bot) = 0$$
$$P(x) = |P(x)|$$
$$P(x \wedge y) \cdot P(y) \cdot P(y)^{-1} = P(x \wedge y) \tag{27}$$

Table 7: WPF_P: weakened axioms for a probability function with name P

structures is denoted by $EPV(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), \mathbb{P})$. $EPV(BA, \mathbb{R}_0(\mathbf{s}), P)$ denotes the union of all collections $EPV(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), P)$ for all $\mathbb{E}$ with $\mathbb{E} \models BA$. It is apparent from the construction that $EPV(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), \mathbb{P}) \models BA + Md + Sign + PF_P$, so that BA+Md+Sign+$\mathrm{PF}_P$ is sound for $EPV(BA, \mathbb{R}_0(\mathbf{s}), P)$. A completeness result for BA+Md+Sign+PF_P is taken from [5].

Theorem 2. *BA+Md+Sign+PF_P is sound and complete for the equational theory of $EPV(BA, \mathbb{R}_0(\mathbf{s}), P)$.*

It is a corollary of the completeness proof in [5] that the same axioms are complete for the class $EPV(BA^f, \mathbb{R}_0(\mathbf{s}), P)$ containing those PF structures which are expansions of a finite event structure.[13]

4.2 BR and BR_2, two forms of Bayes' rule

As a comment to the specification of PFs an excursion to Bayes' rule is meaningful. As a preparation a weaker set of axioms for probability functions is displayed in Table 7. Equation 27 follows from BA+Md+Sign+PF_P. This fact is a consequence of Theorem 2 above and direct proof reads as follows. Let $\phi(u, v) \equiv (1 - \frac{|u| + |v|}{|u| + |v|}) \cdot u$. Now $(\mathbb{R}_0, \mathbf{s}) \models \phi(u, v) = 0$, and using the completeness theorem of [4] one obtains BA+Md+Sign $\vdash \phi(u, v) = 0$. Substituting $P(y \wedge x)$ for u and $P(y \wedge \neg x)$ for v one derives:
BA+Md+Sign+$\mathrm{PF}_P \vdash 0 = (1 - \frac{|P(y \wedge x)| + |P(y \wedge \neg x)|}{|P(y \wedge x)| + |P(y \wedge \neg x)|}) \cdot P(y \wedge x) = (1 - \frac{P(y \wedge x) + P(y \wedge \neg x)}{P(y \wedge x) + P(y \wedge \neg x)}) \cdot P(y \wedge x) = (1 - \frac{P(y)}{P(y)}) \cdot P(y \wedge x)$, from which the required result follows immediately.

Bayes' rule, also known as Bayes' theorem, occurs in different forms. The conditional operator P^0 of Table 6 will be used for its presentation below. The simplest form (hereafter named BR) reads thus:

$$P^0(x \mid y) = \frac{P^0(y \mid x) \cdot P(x)}{P(y)}.$$

In [5] it is shown that BR follows from the specification BA+Md+Sign+PF_P. The short proof makes no use of the axioms in Table 7 only. As it turns out BR implies axiom 27 so that the latter may be viewed as an alternative formulation of BR in the context of

[13] In [11] first order axioms are provided for probability calculus, and corresponding completeness is shown making use of the decidability of the first order theory of reals, a fact which underlies Theorem 2 as well.

BA+Md+Sign+WPF$_P$ This fact is shown as follows: by substituting $x \wedge y$ for y in BR one obtains: $P^0(x|x \wedge y) = (P^0(x \wedge y|x) \cdot P(x))/P(x \wedge y)$. Multiplying both sides with $P(x \wedge y)$ gives $L = R$ with $L = P^0(x|x \wedge y) \cdot P(x \wedge y)$ and $R = ((P^0(x \wedge y|x) \cdot P(x))/P(x \wedge y)) \cdot P(x \wedge y)$. Now $L = (P(x \wedge (x \wedge y))/P(x \wedge y)) \cdot P(x \wedge y) = (P(x \wedge y) \cdot P(x \wedge y))/P(x \wedge y) = P(x \wedge y)$, and $R = (((P((x \wedge y) \wedge x)/P(x)) \cdot P(x))/P(x \wedge y)) \cdot P(x \wedge y) = (P(x \wedge y)/P(x \wedge y)) \cdot P(x \wedge y) \cdot (P(x)/P(x)) = P(x \wedge y) \cdot P(x) \cdot P(x)^{-1}$.

BA+Md+Sign+WPF$_P$ strictly weaker than BA+Md+Sign+PF$_P$. Indeed one may have a four element event space generated by an atomic event e such that $P(\bot) = P(e) = P(\neg e) = 0$ and $P(\top) = 1$ and find that all equations of BA+Md+Sign+WPF$_P$ are satisfied while equation 26 is not satisfied. As BR is provable without the use of axiom 26 it follows that incorporating BR by replacing axiom 26 by BR produces a specification which fails to imply axiom 26.

A second and equally well-known form of Bayes' rule is BR$_2$. BR$_2$ presents a version of Bayes' rule which is used in many educational examples of its application:

$$P^0(x \mid y) = \frac{P^0(y \mid x) \cdot P(x)}{P^0(y \mid z) \cdot P(z) + P^0(y \mid \neg z) \cdot P(\neg z)}.$$

BR, and hence equation 27 follows from BA+Md+Sign+WPF$_P$(minus equation 27)+BR$_2$ by taking $z = \top$.

Moreover, BA+Md+Sign+WPF$_P$+BR$_2$ implies axiom 26. To see this one first notices that, according to [5] it suffices to derive the equation $P(y) = P(y \wedge z) + P(y \wedge \neg z)$, as this equation in combination with BA+Md+Sign+WPF$_P$ entails axiom 26. To this end set $x = y$ in BR$_2$, thereby obtaining $P^0(y|y) = (P^0(y|y) \cdot P(y))/(P^0(y|z) \cdot P(z) + P^0(y|\neg z) \cdot P(\neg z))$.

Using $P^0(y|y) = P(y \wedge y)/P(y) = P(y)/P(y)$ and then taking the inverse at both sides one obtains $L = R$ with $L = P(y)/P(y)$ and $R = (P^0(y|z) \cdot P(z) + P^0(y|\neg z) \cdot P(\neg z))/P(y)$. Then multiplying L and R with $P(y)$ brings $L \cdot P(y) = (P(y)/P(y)) \cdot P(y) = P(y)$ and $R \cdot P(y) =$
$((P^0(y|z) \cdot P(z) + P^0(y|\neg z) \cdot P(\neg z))/P(y)) \cdot P(y) =$
$((P(y \wedge z)/P(z)) \cdot P(z) + (P(y \wedge \neg z)/P(\neg z)) \cdot P(\neg z)) \cdot (P(y)/P(y)) =$
$(P(y \wedge z) + P(y \wedge \neg z)) \cdot (P(y)/P(y)) =$
$P(y \wedge z) \cdot (P(y)/P(y)) + P(y \wedge \neg z) \cdot (P(y)/P(y)) =$
$P(y \wedge z) + P(y \wedge \neg z)$. With $L \cdot P(y) = R \cdot P(y)$ one finds the required result:
$P(y) = P(y \wedge z) + P(y \wedge \neg z)$.

It may be concluded that BR$_2$ provides an adequate substitute axiomeq:26 in the presence of WPF$_P$. This observation suggests an alternative axiomatisation (BA+Md+Sign+PF$'_P$) of probability functions based on BR$_2$. This alternative alternative axiomatisation is presented in Table 8.

4.3 A meadow of conditional values

A third sort named CV involving so-called conditional values is needed. CV is generated by an embedding $\lambda x \in V.v(x)$ and a conditional operator $\lambda e \in E, X \in CV.e :\to X$. CV is equipped with all meadow operations while $v(0)$ serves as 0 and $v(1)$ serves as 1. A specification Md$_{cv}$ of CV is given in Table 9. The conditional operator $\lambda e \in E, X \in CV.e :\to X$ is specified in

$$
\begin{aligned}
P(\top) &= 1 \\
P(\bot) &= 0 \\
P(x) &= |P(x)| \\
P^0(x \mid y) &= P(x \wedge y) \cdot P(y)^{-1} && (28) \\
P^0(x \mid y) &= P^0(y \mid x) \cdot P(x) \cdot (P^0(y \mid z) \cdot P(z) + P^0(y \mid \neg z) \cdot P(\neg z))^{-1} && (29)
\end{aligned}
$$

Table 8: PF'_P: alternative axioms for a probability function with name P.

Table 10.[14] An explicit definition of the *if - then - else* - operator $(x \lhd e \rhd y)$ has been included as a derived operator. It provides no additional expressive power, however, nor does it admit transformation to normal forms which prove useful for the purposes of this paper, and it will not be used below for those reasons.

Given a Boolean algebra $\mathbb{E}$ and a signed meadow $\mathbb{M}(\mathbf{s})$ there is a free and minimal collection $\mathbb{CV}(\mathbb{E}, \mathbb{M}))$ of elements for CV generated from E and $\mathbb{M}(\mathbf{s})$. The three sorted expansion $ECV(\mathbb{E}, \mathbb{M}(\mathbf{s}), \mathbb{CV}(\mathbb{E}, \mathbb{M}(\mathbf{s})))$ of $\mathbb{M}(\mathbf{s})$ and $\mathbb{E}$ includes a sort CV, the conditional operator on E $\times$ CV, and the embedding from V into CV.

Proposition 2. *If $\mathbb{M}$ is nontrivial (that is $\mathbb{M} \not\models 0 = 1$) and $||\mathbb{E}||$ has more than two elements then $\mathbb{CV}(\mathbb{E}, \mathbb{M})$ is not a cancellation meadow (that is $\mathbb{CV}(\mathbb{E}, \mathbb{M}) \not\models X \neq 0 \rightarrow X \cdot X^{-1} = 1$).*

Proof. For each element e of $||\mathbb{E}||$ a constant $\underline{e}$ is introduced, as well as a constant $\underline{m}$ for each element m of $||\mathbb{M}||$. Using the Stone representation of Boolean algebras a set S can be found together with a subset W of the power set of S so that $\mathbb{E}$ is isomorphic to $(W, \cap, \cup, \overline{}, \emptyset, S)$. Let $\phi : E \rightarrow W$ be a homomorphism from $\mathbb{E}$ to the power set algebra of S. To each closed term X of type CV of the extended signature a mapping $[\![X]\!] : S \rightarrow V$ is assigned, with the rules of Table 11.

The equivalence relation $\equiv_S$ on closed CV terms is given by $X \equiv_S Y \iff \forall s \in S([\![X]\!](s) = [\![Y]\!](s))$. $X \equiv_S$ is a congruence relation which meets all requirements imposed by Table 9 and $\mathbb{CV}(\mathbb{E}, \mathbb{M})$ can be defined as the terms of type CV in the extended signature modulo $\equiv_S$.

These preparations enable a proof of the Proposition, by finding an X which differs from $v(0)$ modulo $\equiv_S$ and so that $X \cdot X^{-1}$ differs from $v(1)$ modulo $\equiv_S$. Indeed If $||\mathbb{E}|| > 2$ then an $e \in ||\mathbb{E}||$ exists which is neither $\top$ nor $\bot$, thus $\emptyset \subsetneq \phi(e) \subsetneq S$. Choosing such e it turns out that $e :\rightarrow v(1)$ violates IL. First notice that $\bot :\rightarrow v(1) \not\equiv_S e :\rightarrow v(1) \not\equiv_S \top :\rightarrow v(1)$. Therefore $\bot :\rightarrow v(1) = v(0) \not\equiv_S e :\rightarrow v(1)$. Now $(e :\rightarrow v(1))^{-1} = e :\rightarrow v(1)^{-1} = e :\rightarrow v(1^{-1}) = e :\rightarrow v(1)$

[14] The language of signed meadows can be expanded by introducing a version of the conditional (if-then-else-) operator, which takes an event and two numbers and returns a number. Instead of a three place conditional operator a two place conditional operator will be used which results from the three place conditional by assuming that that the second number argument equals zero.

One may imagine that instead of including an additional sort CV, the conditional values might be viewed as an extension of the sort V. The reason for not doing so lies in the difficulty of finding plausible values for expression of the form $\mathbf{s}(e :\rightarrow x)$. Equation 48 can be simplified to: $e :\rightarrow (X \cdot Y) = (e :\rightarrow X) \cdot Y$ which is equivalent in the presence of the other axioms.

$$v(-x) = -v(x) \tag{30}$$
$$v(x^{-1}) = v(x)^{-1} \tag{31}$$
$$v(x+y) = v(x) + v(y) \tag{32}$$
$$v(x \cdot y) = v(x) \cdot v(y) \tag{33}$$
$$v(x) = v(0) \to x = 0 \tag{34}$$

$$(X+Y)+Z = X+(Y+Z) \tag{35}$$
$$X+Y = Y+X \tag{36}$$
$$X+v(0) = X \tag{37}$$
$$X+(-X) = v(0) \tag{38}$$
$$(X \cdot Y) \cdot Z = X \cdot (Y \cdot Z) \tag{39}$$
$$X \cdot Y = Y \cdot X \tag{40}$$
$$v(1) \cdot X = X \tag{41}$$
$$X \cdot (Y+Z) = X \cdot Y + X \cdot Z \tag{42}$$
$$(X^{-1})^{-1} = X \tag{43}$$
$$X \cdot (X \cdot X^{-1}) = X \tag{44}$$
$$X^2 = X \cdot X$$

Table 9: Md_{cv}: axioms for the meadow of conditional values, $(X, Y, Z$ range over CV)

$$\top :\to X = X \tag{45}$$
$$\bot :\to X = v(0) \tag{46}$$
$$e :\to (X+Y) = (e :\to X) + (e :\to Y) \tag{47}$$
$$e :\to (X \cdot Y) = (e :\to X) \cdot (e :\to Y) \tag{48}$$
$$e :\to (-X) = -(e :\to X) \tag{49}$$
$$e :\to (X^{-1}) = (e :\to X)^{-1} \tag{50}$$
$$(e \vee f :\to X) = (e :\to X) + (f :\to X) - (e \wedge f :\to X) \tag{51}$$
$$e \wedge f :\to X = e :\to (f :\to X) \tag{52}$$
$$X \lhd e \rhd Y = (e :\to X) + (\neg e :\to Y)$$

Table 10: Cond: axioms for the conditional operator

$$[\![v(\underline{m})]\!](s) = m$$
$$[\![-t]\!](s) = -([\![t]\!](s))$$
$$[\![t^{-1}]\!](s) = ([\![t]\!](s))^{-1}$$
$$[\![t+r]\!](s) = [\![t]\!](s) + [\![r]\!](s)$$
$$[\![t \cdot r]\!](s) = [\![t]\!](s) \cdot [\![r]\!](s)$$
$$[\![\underline{e} :\to t]\!](s) = [\![t]\!](s), \textit{if } s \in \phi(e)$$
$$[\![\underline{e} :\to t]\!](s) = 0, \textit{if } s \notin \phi(e).$$

Table 11: Definition of $[\![t]\!](s)$; with $s \in S$

whence $(e :\to v(1)) \cdot (e :\to v(1))^{-1} = (e :\to v(1)) \cdot (e :\to v(1)) = e :\to v(1) \not\equiv_S \top :\to v(1) = v(1)$. □

Definition 3. *An expression $X = e_1 :\to v(t_1) + \ldots e_n :\to v(t_n)$ of type CV is a flat CV expression.*

Definition 4. *A flat CV expression $X = e_1 :\to v(t_1) + \ldots e_n :\to v(t_n)$ is a nonoverlapping flat CV expression if for all $1 \leq i, j \leq n$ with $i \neq j$, it is the case that provably $e_i \wedge e_j = \bot$.*

Definition 5. *Two flat CV expressions are similar if both involve the same collection of conditions.*

Proposition 3. *For each closed CV expression X there is a nonoverlapping flat CV expression Y such that $Md + Sign + Cond \vdash X = Y$.*

Proposition 4. *For all pairs of closed value expressions X and Y similar nonoverlapping flat expressions X' and Y' can be found so that $Md + Sign + Cond \vdash X = X' \,\&\, Y = Y'$.*

Proposition 5. *If we fix $\mathbb{E}$ as some finite minimal algebra then the CVs generated from $\mathbb{E}$ and $\mathbb{R}_0$ constitute a meadow. If $||\mathbb{E}|| > 2$ then the meadow of CVs is not a cancellation meadow.*

Proposition 6. *Given closed CVs in flat form $X = \sum_{i=1}^{n} e_i :\to v(t_i)$ and $Y = \sum_{j=1}^{m} f_j :\to v(r_j)$, a flat form representation for $X \cdot Y$ is: $\sum_{i=1}^{n} \sum_{j=1}^{m} (e_i \wedge f_j) :\to v(t_i \cdot r_j)$. Moreover, if X and Y are in nonoverlapping form so is the given form for $X \cdot Y$.*

4.4 Combining CVs with a PF: expectation values

A CV, say X, denotes a value which is conditional on an event, that is it depends on the actual event e chosen from $\mathbb{E}$. CVs are well-suited suited for defining an expectation value, denoted with $E_P(X)$. The concept of an expectation values lies at the basis of further definitions of probabilistic quantities such as variance, covariance, and correlation. Defining the expectation value for a CV requires a CV in flat form, say $\sum_{i=1}^{n} e_i :\to v(t_i)$, as well as a PF, say P.

$$E_P(\sum_{i=1}^{n} e_i :\to v(x_i)) = \sum_{i=1}^{n} (P(e_i)) \cdot x_i).$$

$$E_P(X+Y) = E_P(X) + E_P(Y) \tag{53}$$
$$E_P(e :\to v(x)) = P(e) \cdot x \tag{54}$$

$$VAR_P(X) = E_P(X^2) - (E_P(X))^2$$
$$COV_P(X,Y) = E_P(X \cdot Y) - E_P(X) \cdot E_P(Y)$$
$$CORR_P^2(X,Y) = \frac{COV_P(X,Y)^2}{VAR_P(X) \cdot VAR_P(Y)}$$

Table 12: EV_P, axioms for the expectation value of a CV relative to PF P

These identities may used as an axiom scheme for the function $E_P : CV \to V$.

Given a PF structure $EPV(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), \mathbb{P})$ and a CV structure $ECV(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), \mathbb{CV}(\mathbb{E}, \mathbb{R}_0(\mathbf{s})))$ a joint expansion $ECVPF(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), \mathbb{CV}(\mathbb{E}, \mathbb{R}_0(\mathbf{s})), \mathbb{P})$ exists. It can be further expanded with E_P to $ECVPF(\mathbb{E}, \mathbb{R}_0(\mathbf{s}), \mathbb{CV}(\mathbb{E}, \mathbb{R}_0(\mathbf{s})), \mathbb{P}, E_P)$.

Instead of using an axiom scheme, a finite axiomatisation of $E_P(-)$ is given in Table 12, from which each instance of the scheme can be derived. The equations (named EV_P) of Table 12 specify $E_P(-)$ on all CVs in the case of a minimal space of conditional values.

In the presence of the equations of EV_P the conditional equation no. 34 of Table 9 above is redundant: indeed assuming $v(x) = v(0)$, $x = P(\top) \cdot x = E_P(\top :\to v(x)) = E_P(\top :\to v(0)) = P(\top) \cdot 0 = 0$.

Grouping together the axioms collected thus far one finds a (conditional) equational theory which: $\mathrm{MbAPC}_P = \mathrm{BA}+\mathrm{Md}+\mathrm{Sign}+\mathrm{PF}_P+\mathrm{Md}_{cv}+\mathrm{Cond}+\mathrm{EV}_P$.[15] A plausible class of models for MbAPC_P is $ECVPF(BA, \mathbb{R}_0(\mathbf{s}), \mathbb{CV}(BA, \mathbb{R}_0(\mathbf{s})), P, E_P)$. Here it is assumed that An event space $\mathbb{E} \models BA$ is chosen and CVs are generated from $\mathbb{E}$ and $\mathbb{R}_0(\mathbf{s})$.

It is easy to see that E_P can be eliminated from expressions of sort V with out free variables of sort CV. Therefore an equation of type V without free variables of sort CV V is provably equivalent within MbAPC_P to an equation not involving subterms of sort CVs. Using Theorem 2, it follows that MbAPC_P is complete for such equations w.r.t. validity in $ECVPF(BA, \mathbb{R}_0(\mathbf{s}), \mathbb{CV}(BA, \mathbb{R}_0(\mathbf{s})), P, E_P)$.

4.5 Variance, covariance, and correlation for conditional values

On the basis of a definition of expectation value, variance, covariance, and correlation on CVs can be introduced as derived operators.

Let X and Y be CVs with flat forms $X = \sum_{i=1}^{n} e_i :\to v(t_i)$ and $Y = \sum_{i=1}^{m} f_i :\to v(r_i)$. The equations in Table 12 provide explicit definitions of variance, covariance, and correlation for X, resp. Y.

There is no novelty to these definitions except for the effort made to make each definition

[15]Here MbAPC stands for meadow based axiomatic probability calculus.

fit a framework that has been setup on the basis of an algebraic specification. By proceeding in this manner an axiomatic framework is obtained for equational reasoning about each of these technical notions.

Leaving out out the subscript, that is using $E(X)$ instead of $E_P(X)$, and similarly for the other operators, is common practice in probability theory. Doing so, however requires that it is apparent from the context which probability function is used. Moreover it must be assumed that for X and for Y the same probability function is used, and for that reason that X and Y share the same sample space.

4.6 Extracting a PMF from a CV

Given a CV X with nonoverlapping flat form $\sum_{i=1}^{n} e_i :\to v(y_i)$, and a probability function P the probability mass function, $\lambda x.P(X = x)$ for X is supposed to yield for each value x the probability that X takes value x.[16] An explicit definition for the PMF of X is as follows:

$$P((\sum_{i=1}^{n} e_i :\to v(y_i)) = x) = \sum_{i=1}^{n} (0_{y_i - x} \cdot P(e_i)).$$

Proposition 7. *Equivalence of definitions for expectation value and variance for CV expressions in nonoverlapping flat form via (joint) PMFs extraction.*

1. $E_P(X) = E_{pmf}(\lambda x.P(X = x))$,
2. $VAR_P(X) = VAR_{pmf}(\lambda x.P(X = x))$.

Proof. Let $X = \sum_{i=1}^{n} e_i :\to y_{)}$ be a nonoverlapping flat CV expression. Making use of the facts listed in Paragraph 2.1, one obtains: $E(\lambda x.P(X = x)) = \sum_x^\star \sum_{i=1}^{n} (0_{y_i - x} \cdot P(e_i)) = \sum_{i=1}^{n} \sum_x^\star (0_{y_i - x} \cdot P(e_i)) = \sum_{i=1}^{n} \sum_x^\star (0_{y_i - x} \cdot P(e_i)) = \sum_{i=1}^{n} (y_i \cdot P(e_i)) = E_P(X)$. □

4.7 Joint PMF extraction for event sharing CVs

Two CVs are event sharing if both have conditions over the same domain. Extraction of a joint PMF from event sharing CVs works as follows. Given two CVs X and Y with similar nonoverlapping flat forms $\sum_{i=1}^{n} (e_i :\to y_i)$ and $\sum_{i=1}^{n} (e_i :\to z_i)$ the joint PMF for these CVs, conventionally written as $P(X = x, Y = y)$, is defined by

$$P(X = x, Y = y) = \sum_{i=1}^{n} (0_{y_i - x} \cdot 0_{z_i - y} \cdot P(e_i)).$$

Extending Proposition 4.6 the following connections between definitions acting on a CV and definitions acting on a PMF or on a joint PMF can be found.

[16] The notation $P(X = x)$ in case X is a CV reflects the corresponding notational convention for PMF extraction from a random variable.

Proposition 8. *Equivalence of definitions for covariance and correlation (squared) via CVs and via (joint) PMFs.*

1. $COV_P(X,Y) = COV_{pmf}(\lambda x, y.P(X = x, Y = y))$,
2. $CORR^2_P(X,Y) = CORR^2_{pmf}(\lambda x, y.P(X = x, Y = y))$.

4.8 Independence of CVs

Independence of CVs can be defined with the help of joint PMF extraction. Two CVs X and Y are independent w.r.t. the probability function P, if for all $x, y \in \mathbb{R}_0$: $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$.

5 Configuration space

A configuration space (CS) allows the specification of structures and constellations about which probabilistic reasoning may take place. A motive for the introduction of the sort CS is to enable the formalisation of informal ways of speaking of agents who are capable of maintaining and revealing preferences about different configurations.

5.1 Signature and axioms for configuration spaces

A configuration is a parallel composition of objects each of which may be conditionally present. In addition each object, and in more generally each configuration, may carry with it a yield in CV which expresses its value. The following constructors for configurations are used:

1. The finite set *Obj* contains constants for objects, which may or may not exist in parallel in a world of states. These constants point to the primitive objects from which configurations are made up.
2. The empty state ϵ, (thus following the notational conventions of [13]),
3. Parallel composition for states α and β, written as $\alpha \| \beta$.
4. Conditional states, written $e :\rightarrow \alpha$, for an event e and a state α.
5. Attaching a CV X to a state α, written $\alpha \Rsh X$. In a configuration $\beta = \alpha \Rsh X$, X is the CV carried by β.

Equations for parallel object states are listed in Table 13. Moreover a utility function from CS to CV is specified in Table 14. Given an expectation value operator on CV, an expected utility value can be defined on POS, as indicated in Table 14. Taking axioms together: Md + BA + Md_{cv} + Cond + CS provides axioms for events, values, conditional values, and configurations. Extending with PF + UF takes probability functions into account.

$$\alpha \| \beta = \beta \| \alpha \tag{55}$$
$$(\alpha \| \beta) \| \gamma = \beta \| (\alpha \| \gamma) \tag{56}$$
$$\alpha \| \epsilon = \alpha \tag{57}$$
$$\top :\rightarrow \alpha = \alpha \tag{58}$$
$$\bot :\rightarrow \alpha = \epsilon \tag{59}$$
$$e :\rightarrow (\alpha \| \beta) = (e :\rightarrow \alpha) \| (e :\rightarrow \beta) \tag{60}$$
$$(e :\rightarrow \alpha) \| (f :\rightarrow \alpha) = ((e \vee f) :\rightarrow \alpha) \| ((e \wedge f) :\rightarrow \alpha) \tag{61}$$
$$c = c \Rsh v(0) \quad \text{for c in Obj} \tag{62}$$
$$(\alpha \Rsh X) \Rsh Y = \alpha \Rsh Y \tag{63}$$
$$\epsilon \Rsh X = \epsilon \tag{64}$$
$$(\alpha \| \beta) \Rsh X = (\alpha \Rsh X) \| (\beta \Rsh X) \tag{65}$$

Table 13: CS: equations for configuration space

5.2 Comparison of configurations I: absence of preference

The setup with sorts CV and CS relates in a plausible manner with classical texts on subjective probabilities. Two examples of this connection will be discussed, both able of being modeled by means of parallel object states.

Suppose that u_1 and u_2 are closed terms for signed meadows which differ in $\mathbb{R}_0(\mathbf{s})$. A PF named P^A is assumed to quantify the subjective probability that an agent A assigns to various events. It is assumed that A can be questioned about its preference between different states, presented to it as options between which it may freely choose, where it is supposed to prefer states with a higher expected utility.

Let $option_1$ and $option_2$ be chosen as follows: $option_1 = (e :\rightarrow c_1 \Rsh u_1) || (\neg e :\rightarrow c_2 \Rsh u_2)$ and $option_2 = (e :\rightarrow c_2 \Rsh u_2) || (\neg e :\rightarrow c_1 \Rsh u_1)$, and that $option_1$ and $option_2$ have equal preference according to A, then it may be assumed that $P^A(e) = 1/2$. In this manner subjective probabilities may be uncovered from information about utilities and preferences.

Now consider two options as follows: $option_1 = (e :\rightarrow c_1 \Rsh u_1) || (\neg e :\rightarrow c_2 \Rsh u_2)$ and $option_3 = (e :\rightarrow c_3 \Rsh u_3) || (\neg e :\rightarrow c_4 \Rsh u_4)$ and in addition assume that $u_4 - u_2 + u_1 - u_3 \neq 0$. It is further assumed that A has no preference for either option above the other option. Then A's indifference between both options implies $E^U_{P^A}(option_1) = E^U_{P^A}(option_2)$ which reveals a subjective probability $P^A(e)$ given by:

$$P^A(e) = \frac{u_4 - u_2}{u_4 - u_2 + u_1 - u_3}.$$

These observations merely constitute reformulations of observations made in [10] where the underlying idea was credited to [16].

$$
\begin{aligned}
U(\epsilon) &= v(0) && (66)\\
U(c \upharpoonright X) &= X, (c \in Obj) && (67)\\
U(\alpha \| \beta) &= U(\alpha) + U(\beta) && (68)\\
U(e :\rightarrow \alpha) &= e :\rightarrow U(\alpha) && (69)\\
E_P^U(\alpha) &= E_P(U(\alpha))
\end{aligned}
$$

Table 14: UF: equations for a utility function

5.3 Comparison of configurations II: disutility of asking

In [16] one finds a setup which can be formalised as follows: Agent A has made a choice to take either of two paths, there were no other alternatives available to A. The presence of state component π_1 represents A having made the choice to follow the first path, and π_1 indicates A's having made the other choice. Depending on event e, the outcome of which is unknown to A, either π_1 or π_2 produces outcome $v(high)$ and the other path produces $v(low)$, with $high$ and low values in $\|\mathbb{R}_0\|$ so that $low < high$. A prefers to find a high yield rather than the low one; agent A will find the yield produced by following a path just by not planning any move to leave the path.

However, A is capable of asking the outcome of e at cost $d \in V$, and to change to the other path when knowing the outcome of e suggests a higher return. The assumption is that at cost d (disutility d) A is willing to ask about e and to change path when needed. This state of affairs is made more formal by defining two options, $option_1$ and $option_2(d)$ the second option depending on a parameter d ranging over V, and asserting that A prefers $option_2(d)$ over $option_1$. Both options are specified in CS as follows:

$option_1 = \pi_1 \upharpoonright (e :\rightarrow v(high) + \neg e :\rightarrow v(low))$

$option_2(d) = e :\rightarrow \pi_1 \upharpoonright (v(high - d) + \neg e :\rightarrow \pi_2 \upharpoonright v(high - d))$.

The asserted preference that A maintains for $option_2(d)$ over $option_1$, that is A's willingness to ask about e (at cost d) and to change path when needed, is equivalent to $E_{P^A}^U(option_1) < E_{P^A}^U(option_1)$. After some calculation it appears that A produce said preference if

$$
P^A(e) < 1 - \frac{d}{high - low}.
$$

It follows that by observing A's preferences about pairs of options $option_1$ and $option_2(d)$, an observer of A can determine upper bounds for $P^A(e)$. If a lowest upper bound is determined that information precisely determines $P^A(e)$ as well.

6 The multidimensional case

In the multidimensional case event space is considered a product of event spaces, with a dedicated event space for different CVs. In the multi-dimensional case CVs occurring in a vector of CVs are by default not event space sharing and the notion of a joint probability function working over a tuple of event spaces enters the picture.

The multi-dimensional case becomes relevant once tuples (vectors) of CVs are considered in combination with a plurality of joint PFs for product spaces of higher dimensional event space corresponding to various vectors of CVs such that there may not exist a joint PF for the full product space.[17]

6.1 Multidimensional probability functions

Let $D = \{a_1, \ldots, a_d\}$ be a finite set. The elements of D will be called dimensions, D is called a dimension set.

Definition 6. *(Arities over D) ar_D, the collection of arities over dimension set D, denotes the set of finite non-empty sequences of elements of D without repetition.*

Elements of ar_D will serve as arities of probability functions on multi-dimensional event spaces. $l(w)$ denotes the length of $w \in ar_D$.

Definition 7. *(Arity family) Given an event space E, and a name P for a probability function, an arity family (for E and P) is a finite subset W of ar_D which is (i) closed under permutation, and (ii) closed under taking non-empty subsequences, and (iii) which contains for each $d \in D$ the arity (d), that is the one-dimensional arity consisting of dimension d only.*

The presence of a function $P_w : E^{l(w)} \to V$ for each $w \in W$ is assumed. Instead of axiomatising a single probability function we will now provide axioms for a probability function family (see Table 15). Because in an arity repetition of dimensions is disallowed these axioms reduce to what we had already in the case of a single dimension.

The generalisation to the multidimensional case of the axiom system BA+Md+Sign+MD$_{cv}$+Cond+PF$_P$ reads BA+Md+Sign+MD$_{cv}$+Cond+PFF$_{W,P}$.

6.2 Multivariate conditional values

A multivariate CV is a vector of CVs where each component is labeled with a unique dimension taken from a finite set D of dimensions. In the context of an arity family W it is assumed that the corresponding vector of labels is in W. A multivariate CV for W is a tuple $\langle X_1^{d_1}, \ldots, X_n^{d_n} \rangle$ with X_i CVs and with $\langle d_1, \ldots, d_n \rangle \in W$. The different component of a multivariate CVs are not event space sharing, though the different event spaces may be plausibly chosen as isomorphic copies, a choice which will be adopted below and which also underlies Table 15.

[17] Precisely this situation is what the Bell inequalities are about. For more details see [5]. The terminology of conventional probability theory, by being set up on the basis of a single sample space and the corresponding event space, seems to be biased to the case that joint probability distributions for on products of event spaces always exist.

$$P^{d,u,e,u'}(y_1, x_1 \ldots, x_l, y_2, z_1, \ldots, z_{l'}) = P^{d,u,e,u'}(x_2, x_1 \ldots, x_l, y_1, z_1, \ldots, z_{l'}) \quad (70)$$
$$P^d(\top) = 1 \quad (71)$$
$$P^d(\bot) = 0 \quad (72)$$
$$P^{d,w}(\top, x_1, \ldots, x_n) = P^w(x_1, \ldots, x_n) \quad (73)$$
$$P^{d,w}(\bot, x_1, \ldots, x_n) = 0 \quad (74)$$
$$P^w(x_1, \ldots, x_n) = |P^w(x_1, \ldots, x_n)| \quad (75)$$
$$P^{d,u}(x \vee y, x_1, \ldots, x_l) = P^{d,u}(x, x_1, \ldots, x_l) + P^{d,u}(y, x_1, \ldots, x_l) - P^{d,u}(x \wedge y, x_1, \ldots, x_l) \quad (76)$$

Table 15: $\mathrm{PFF}_{W,P}$: axioms for a probability function family with name P (with $d, e \in D$, $w, (d, u), (c, u, d, u') \in W, n = l(w)$, and $u, u' \in ar_D \cup \{\epsilon\}, l = l(u), l' = l(u')$.

By taking $D = \{a, b\}$ one obtains the case the case with two dimensions which is relevant for the definitions of covariance and correlation. In this case four forms of multivariate CVs are encountered: $\langle X^a \rangle$ and $\langle Y^b \rangle$ as univariate CVs , and bivariate (that is 2-dimensional) CVs of the form $\langle X^a, Y^b \rangle$ as well as of the form $\langle X^b, Y^a \rangle$.

6.3 Expectation, (co)variance, and correlation, the multivariate case

Given a CV X with nonoverlapping flat form representation $X = \sum_{i=1}^{k} e_i :\rightarrow v(t_i)$, and a CV Y with flat form representation $t = \sum_{i=1}^{k} e_i :\rightarrow v(r_i)$ and a PFF with name P and arity family W with $(a, b) \in W$ the definitions of expectation value, variance, covariance, and correlation, can be detailed as follows.

Expectation value:

$$E_P(\langle X^a \rangle) = E_{P^a}(X) = \sum_{i=1}^{k} t_i \cdot P^a(e_i)$$

Variance:

$$VAR_P(\langle X^a \rangle) = VAR_{P^a}(X) = \sum_{i=1}^{k} (t_i^2 \cdot P^a(e_i)) - (\sum_{i=1}^{k} t_i \cdot P^a(e_i))^2$$

Covariance:

$$COV_P(\langle X^a, Y^b \rangle) = \sum_{i=1}^{k} (\sum_{j=1}^{k} (t_i \cdot r_j \cdot P^{a,b}(e_i, e_j))) - E_P(\langle X^a \rangle) \cdot E_P(\langle Y^b \rangle)$$

Correlation:

$$CORR_P^2(\langle X^a, Y^b \rangle) = \frac{COV_P(\langle X^a, Y^b \rangle)^2}{VAR_P(\langle X^a \rangle) \cdot VAR_P(\langle Y^b \rangle)}$$

$$\top = \langle \top, \top \rangle \tag{77}$$
$$\bot = \langle e, \bot \rangle \tag{78}$$
$$\bot = \langle \bot, e \rangle \tag{79}$$
$$\langle e_1, e_2 \rangle \wedge \langle f_1, f_2 \rangle = \langle e_1 \wedge f_1, e_2 \wedge f_2 \rangle \tag{80}$$
$$\langle e, f \vee g \rangle = \langle e, f \rangle \vee \langle e, g \rangle \tag{81}$$
$$\langle e \vee f, g \rangle = \langle e, g \rangle \vee \langle f, g \rangle \tag{82}$$

Table 16: additional axioms for Boolean operations in 2 dimensions

6.4 Reduction to the 1-dimensional case

The multidimensional case is primarily interesting in case the arity family W has no maximal element. If W has a largest element $u_1, \ldots, u_n$ then the PFF can be identified with $P^{u_1,\ldots,u_n}$ understood as a 1-dimensional PF over an event space with as its domain the cartesian product of the event spaces $\mathbb{E}^{u_1}, \ldots, \mathbb{E}^{u_n}$. The details of this transformation are presented in the 2-dimensional case only, that case being of particular relevance for the definitions of covariance and of correlation.

For 2 dimensions, corresponding with the dimensions say a and b of D, a new sort $E^{a,b}$ is introduced with elements of the form $\langle e, f \rangle$ with $e, f \in \mathbb{E}$, and the signature of BA is copied for that sort. In addition to these operations the 2-place conditional operator from events and values to values is introduced on sort $E^{a,b}$. The axioms of Table 16 in addition to BA turn the new sort into a Boolean algebra which is properly connected with *events*. Closed conditional expressions over $E^{a,b}$ can be written in flat form.

A probability function $P^{a \times b}$ is determined by the equation in Table 18 in addition to the axioms for probability functions stated for $P^{a \times b}$. Besides a new sort for events also a new sort $\mathrm{CV}^{a,b}$ for CVs is needed. CV expressions X over E are lifted to CV expressions in $\mathrm{CV}^{a,b}$. The transformation $\lambda X \in CVE.[X]_0$ is applied to CVs for dimension a, and $\lambda X \in CVE.[X]_1$is applied to CVs for dimension b (where CVE stands for CV expressions). The operation of $[X]_\alpha$ for both values $\alpha \in \{0, 1\}$ is specified in Table 17.

The definitions of expectation value, variance, covariance, and correlation squared can alternatively be given by translating the 2 dimensional case back to a 1-dimensional case. These explicit definitions are given in Table 19.

7 Concluding remarks

This paper is a sequel to [5] where a meadow based approach to the equational specification of probability functions was proposed. The equational axiomatic presentation of probability functions has been extended with a conditional operator which generates so-called conditional values (CVs), that is values conditional on an event, and an expectation value operator. On top of these axiomatic definitions variance is defined for a single conditional value and covariance

$$[v(x)]_\alpha = v(x) \tag{83}$$
$$[-X]_\alpha = -[X]_\alpha \tag{84}$$
$$[X^{-1}]_\alpha = ([X]_\alpha)^{-1} \tag{85}$$
$$[X+Y]_\alpha = [X]_\alpha + [Y]_\alpha \tag{86}$$
$$[X \cdot Y]_\alpha = [X]_\alpha \cdot [Y]_\alpha \tag{87}$$
$$[g :\rightarrow X]_0 = \langle g, \top \rangle :\rightarrow [X]_a \tag{88}$$
$$[g :\rightarrow X]_1 = \langle \top, g \rangle :\rightarrow [X]^b \tag{89}$$

Table 17: Lifting CVs to $\mathrm{CV}^{a,b}$.

$$P^{a\times b}(\langle e, f\rangle) = P^{a,b}(e, f) \tag{90}$$

Table 18: Lifting a PFF to product space.

$$\begin{aligned}
E_{P^{a\times b}}([X]_0)) &= E_{P^a}(X)\\
E_P(\langle X^a\rangle) &= E_{P^{a\times b}}([X]_0)\\
VAR_P(\langle X^a\rangle) &= E_{P^{a\times b}}([X^2]_0) - E_{P^{a\times b}}([X]_0)^2\\
COV_P(\langle X^a, Y^b\rangle) &= E_{P^{a\times b}}([X]_0 \cdot [Y]_1) - E_{P^{a\times b}}([X]_0) \cdot E_{P^{a\times b}}([Y]_1),\\
CORR_P^2(\langle X^a, Y^b\rangle) &= \frac{COV_P(\langle X^a, Y^b\rangle)^2}{VAR_P(\langle X^a\rangle) \cdot VAR_P(\langle Y^b\rangle)}.
\end{aligned}$$

Table 19: Derived operators after reduction to the 1-dimensional case

and correlation are defined for pairs of conditional values sharing the same event space. An alternative development is provided in the higher dimensional setting where conditional values are not supposed to share the respective event spaces and where situations can be formally modeled in which joint probability functions are not guaranteed to exist.

The notion of a CV plays the role of a discrete random variable with the additional restriction that it takes finitely many values only. By working with CVs the use of a sample space underlying the event space is avoided thus allowing to maintain the style, and to a lesser extent, the simplicity, of the axiomatisations of probability functions that have been proposed in [5]. On the basis of the sort CV of CVs, a configuration space CS is built which allows formalising more involved contexts.

As a technical tool finite support summation (FSS) is introduced, a novel binding operator on meadows. FSS is of independent interest for the theory of meadows and it gives rise to intriguing new questions. Using FSS a concise specification of the class of finitely supported probability mass functions (PMFs) is provided independently of the notion of a probability function.

The development of meadow based axiomatic probability calculus makes use of zero as the result of dividing by zero, and does so without hesitation. This particular design decision is helpful for obtaining equational specifications. The scope of equational axiomatisations is limited. For instance proving the Dutch Book Theorem requires working with existential quantification over probability functions which is not supported by equational reasoning. It is worth mentioning that applications of this particular way of dealing with division by zero are currently being proposed in other areas of elementary mathematics as well, for instance in matrix theory (see [14]).

Acknowedgement: Yoram Hirschfeld, Kees Middelburg and Alban Ponse have provided several useful comments on previous versions of the paper.

References

[1] S. Abramsky. Relational databases and Bell's theorem. In: *In search of elegance in the theory and practice of computation.* Springer Berlin Heidelberg, 13-35, 2013.

[2] D. Barber. *Bayesian Reasoning and Machine Learning.* Cambridge University Press, 2012. (ISBN 0521518148, 9780521518147). On-line version available at `http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.Online` (consulted version: 18 June 2013).

[3] J.A. Bergstra, I. Bethke, and A. Ponse. Cancellation meadows: a generic basis theorem and some applications. *The Computer Journal*, 56(1):3–14, 2013. (`doi:10.1093/comjnl/bxs028`). Also available at `http://arxiv.org/abs/0803.3969`.

[4] J.A. Bergstra, I. Bethke, and A. Ponse. Equations for formally real meadows. *Journal of Applied Logic*, 13(2) part B:1–23, 2015. (`doi:10.1016/j.jal.2015.01.004`). Also available at `http://arxiv.org/abs/1310.5011`.

[5] J.A. Bergstra and A. Ponse. Signed meadow valued probability functions. http://arxiv.org/abs/1307.5173v2, 2016.

[6] J.A. Bergstra and C.A. Middelburg. Inversive meadows and divisive meadows. *Journal of Applied Logic*, 9(3): 203–220, 2011. Also available at http://arxiv.org/abs/0907.0540.

[7] J.A. Bergstra and C.A. Middelburg. Probabilistic thread algebra. *SACS*, 25(2): 211–243, 2015. Also available at http://arxiv.org/abs/1409.6873.

[8] J. A. Bergstra and J. V. Tucker. The rational numbers as an abstract data type. *J. ACM*, 54, 2, Article 7 (April 2007) 25 pages, 2007.

[9] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability*, Athena Scientific, Nashua USA, ISBN 978-1-886529-23-6, 2008.

[10] D. Davidson and P. Suppes. A Finitistic Axiomatization of Subjective Probability and Utility. Econometrica 24 (3) 264-275, 1956.

[11] J.Y. Halpern. An analyisis of first-order logics of probability. Artifician Inteligence 46, 311-350, 1990.

[12] Khan Academy. Random variables and probability distributions. https://www.khanacademy.org/math/probability/random-variables-topic/random-variables-prob-dist/v/discrete-and-continuous-random-variables, consulted July 9 2016.

[13] C.P.J. Koymans, and J. L. M. Vrancken. Extending process algebra with the empty process. Electronic, report LGPS 1. Dept. of Philosophy, State University of Utrecht, The Netherlands (1985).

[14] T. Matsuura and S. Saitoh. Matrices and Division by Zero. Advances in Linear Algebra & Matrix Theory 6: 51-58 (http:dx.doi.org/10.4236/alamt.2016.62007), 2016.

[15] H. Padmanabhan. A self-dual equational basis for Boolean algebras. *Canad. Math. Bull.*, 26(1):9–12, 1983.

[16] F.P. Ramsey. Truth and Probability (1926). in Ramsey, (1931) *The Foundations of Mathematics and other Logical Essays* Ch. VIII, 199-211, edited by R.B. Braithwaite, London, New York, (1999 electronic edition).

[17] G. Shafer and V. Vovk. The Sources of Komogorov's *Grundbegriffe*. Statistical Science, 21 (1) 70-98, 2006.

[18] Wikipedia. https://en.wikipedia.org/wiki/Random_variable, consulted July 9, 2016.

A Random variables

The notion of a random variable (RV) plays a central role in many presentations of probability theory. In the presentation of the current paper the role of a RV is played by a CV instead. In this Appendix it will be outlined how to view a CV as an RV under the additional assumption that event space is finite.

A.1 Implicit sample space

An event $e \in ||\mathbb{E}||$ is atomic if for each event $f \in ||\mathbb{E}||$ either $e \wedge f = e$ or $e \wedge f = \bot$. A finite event space has a finite number of atomic events, say $a_1, \ldots, a_n$ which are assumed to be pairwise different. Given an enumeration without repetition of atomic events, each event e can be written in a unique way as $a_{i_1} \vee \ldots \vee a_{i_k}$ for appropriate $1 \leq i_1 < ... < i_k \leq n$. The collection of atomic events is denoted AE.

The idea is to view the event space in terms of its Stone representation, as has been done in the proof of Proposition 2 above. The notation from that proof will be used below. With the additional constraint that $||\mathbb{E}||$ is finite the set S can be taken to be the set of atomic events so that each event can be identified with a finite set of atomic events (AEs). AE can be considered the sample space that underlies $\mathbb{E}$.[18] The collection AE may alternatively be referred to as the implicit sample space of $\mathbb{E}$.

An RV is a function from sample space to values. Every CV expression X determines a function $[\![X]\!]$ from S to V. The relation between an expression an the graph of its associated function is given by the following Proposition.

Proposition 9. *Let $\sum_{i=1}^{n} e_i :\rightarrow v(t_i)$ be a nonoverlapping flat form for X, and let $a \in AE$ so that $e_j \wedge a = a$ for some j with $1 \leq j \leq n$. Then : $[\![X]\!](a) = t_i$.*

A dedicated binding operator for summation is needed for the formalisation of requirements connected with atomic events. Summation over atomic events, (working correctly if there are only finitely atomic events), is specified as follows:

Definition 8. *Given an event space $\mathbb{E}$ and a term t of sort V, then $\sum_{\alpha \in AE}^{\star} t = 0$ if there are either none or infinitely many atomic events in $||\mathbb{E}||$ and otherwise $\sum_{\alpha \in AE}^{\star} t = \sum_{i=1}^{k} [a_i/\alpha]t$ with $a_1, \ldots, a_k$ an enumeration without repetitions of the atomic events of $\mathbb{E}$.*

A.2 The expectation value of an RV

By regarding AE as a sample space an additional constraint on PFs arises: $\sum_{\alpha \in AE}^{\star} P(\alpha) = 1$. The validity of this assumption is guaranteed by the axioms for a PF if $\mathbb{E}$ is finite.

[18] I prefer not to have samples (that is AE) as a sort because that leads to the setting of subsorts which is not so easy to reconcile with equational logic. Such difficulties persist in spite of the presence of many works that have been devoted to that particular issue. It should be noted that in case subsorts are considered unproblematic (in connection with equational logic) partial functions should be considered harmless as well. Indeed by considering the non-zero reals a subsort of the reals the partial inverse can be appropriately typed. In other words, in a setting of equational logic with subsets the use of meadows as a substitute for working with a partial inverse operator is unconvincing.

Let X be a closed CV expression. Then, in accordance with Table 11, $[\![X]\!]$ is a function from AE to V. Viewing AE as a sample space $[\![X]\!]$ qualifies as an RV. Under the aforementioned assumption that $\sum^{\star}_{\alpha \in AE} P(\alpha) = 1$ the expectation value of the RV $[\![X]\!]$ is defined in the conventional way by:

$$E_P([\![X]\!]) = \sum^{\star}_{\alpha \in AE} ([\![X]\!](\alpha) \cdot P(\alpha)).$$

It follows from this definition that $E_P([\![X]\!] \cdot [\![Y]\!]) = \sum^{\star}_{\alpha \in AE}([\![X]\!](\alpha) \cdot [\![Y]\!](\alpha) \cdot P(\alpha))$. Definitions of variance, covariance, and correlation (squared) are standard as well.

Variance:

$$VAR_P([\![X]\!]) = E_P([\![X]\!]^2) - E_P([\![X]\!])^2$$

Covariance:

$$COV_P([\![X]\!], [\![Y]\!]) = E_P([\![X]\!] \cdot [\![Y]\!]) - E_P([\![X]\!]) \cdot E_P([\![Y]\!])$$

Correlation:

$$CORR^2_P([\![X]\!], [\![Y]\!]) = \frac{COV_P([\![X]\!], [\![Y]\!])^2}{VAR_P([\![X]\!]) \cdot VAR_P([\![Y]\!])}$$

A.3 RVs in colloquial language

RVs play a key role in many accounts of probability theory. However, the concept of a random variable seems to be rather informal and its use is often cast in colloquial language. A common wording states that

> a random variable is the outcome of a stochastic process.[19]

In [18] one reads about a random variable that it is:

> ... a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense).... A random variable can take on a set of possible different values (similarly to other mathematical variables), each with an associated probability, in contrast to other mathematical variables.'

In [12] an RV is explained as a mapping from "outcomes" to values which provides quantification, while the main argument put forward for the introduction of an RV is about the use of its name, and at the same time the suggestion is made that an RV is linked to a probability function.[20] In [9] p. it is stated that

[19] As a definition this is somehow overdetermined, perhaps comparable the hypothetical definition of a real number as "a real number is the length of a ship". Better is: "a random variable is an outcome of a stochastic process", and even better is: "a random variable, for a stochastic process, is a numerical outcome of that same process". The latter assertion is valid, but it fails to capture the independence of the notion of an RV from any specific stochastic process.

[20] The relation between an RV and a PF is like the relation between a car and a driver. In colloquial language a car has a driver but no reference is made to any specific specific person playing that role, and a car can be grasped separately from its driver. Whenever the car is used "as a car", however, it has a unique driver.

> A discrete random variable has an associated probability mass function (PMF),..

In the introductory probability refresher of [2] the domain of a variable is said to be the set of states it can take, while relation between (random) variables and events is explained as follows:

> For our purposes, events are expressions about random variables, such as *Two heads in 6 coin tosses*.

Explanatory statements as mentioned above may serve to produce an intuition for the notion of an RV. What makes understanding RVs difficult, however, is that the mathematical definition of an RV, which reads "a function from sample space to reals" makes no reference to any variable or variable name, or to a probability function. Perhaps one must read "variable" in RV as "function" with an emphasis on variable output rather than on variable input (as in function of one variable), and "random" as "occurring in probability theory".

Under the condition of a finite event space the phrase *conditional value* seems to capture the idea behind an RV quite well, and it is less prone to misunderstandings than the phrase *random variable*.

A.4 Conventional rationale of the notion of an RV

When throwing a dice the outcome is an orientation of the dice which is characterised by a pattern. Calculating with such an outcome is difficult, and if when done it may be ad hoc. The definition of an expectation value, however, requires the presence of addition and multiplication, as well as the availability of a probability function. The simplest examples of probability functions require the use rational fractions. These observations suggests that values are to be taken from a structure which is equipped with the signature of a meadow at least.

After converting outcomes of a probabilistic process to values in a number system, for instance a meadow, well-known quantified algebra, or calculus depending on the number system of choice, can be applied. Enabling calculations with process outcomes this is precisely what an RV achieves by converting a (possibly non-numerical) variable outcome of a (probabilistic) process to a (numerical) value.

From the viewpoint of computer science, however, more can be said about the logic of RVs, for instance the presentation of basics of probability theory in [1] uses a categorial approach and subsequently indicates that the theory of RVs together with probability functions can be profitably developed in close correspondence to the theory of relational database schemas.

In the present paper working with CVs instead of RVs allows in an initial phase of theory development, to take event space as the logical point of departure rather than to have sample space play that role.

Contents