Predictive inference under exchangeability, and the Imprecise Dirichlet Multinomial Model

Gert de Cooman, Jasper De Bock, and Márcio Diniz

Abstract Coherent reasoning under uncertainty can be represented in a very general manner by coherent sets of desirable gambles. In this framework, and for a given finite category set, coherent predictive inference under exchangeability can be represented using Bernstein coherent cones of multivariate polynomials on the simplex generated by this category set. This is a powerful generalisation of de Finetti's Representation Theorem allowing for both imprecision and indecision. We define an inference system as a map that associates a Bernstein coherent cone of polynomials with every finite category set. Many inference principles encountered in the literature can then be interpreted, and represented mathematically, as restrictions on such maps. We discuss two important inference principles: representation insensitivity—a strengthened version of Walley's representation invariance—and specificity. We show that there is a infinity of inference systems that satisfy these two principles, amongst which we discuss in particular the inference systems corresponding to (a modified version of) Walley and Bernard's Imprecise Dirichlet Multinomial Models (IDMMs) and the Haldane inference system.

1 Introduction

This paper deals with predictive inference for categorical variables. We are therefore concerned with a (possibly infinite) sequence of variables X_n that assume values in some finite set of categories A. After having observed a number $\check{n}$ of them, and having found that, say $X_1 = x_1$, $X_2 = x_2$, ..., $X_{\check{n}} = x_{\check{n}}$, we consider some subject's belief model for the next $\hat{n}$ variables $X_{\check{n}+1}, \dots X_{\check{n}+\hat{n}}$. In the probabilistic tradition—and we want to build on this tradition in the context of this paper—this belief can be modelled by some conditional predictive probability mass function $p^{\hat{n}}(\cdot|x_1,\dots,x_{\check{n}})$ on the set $A^{\hat{n}}$ of possible values for these next variables. These probability mass functions can

Ghent University, SYSTeMS Research Group, Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium, e-mail: {gert.decooman,jasper.debock,marcio.diniz}@UGent.be

be used for prediction or estimation, for statistical inferences, and in decision making involving the uncertain values of these variables. In this sense, predictive inference lies at the heart of statistics, and of learning under uncertainty.

What connects these predictive probability mass functions for various values of $\check{n}$, $\hat{n}$ and $(x_1,\dots,x_{\check{n}})$ are the requirements of *temporal consistency* and *coherence*. The former requires that when $n_1 \leq n_2$, $p^{n_1}(\cdot|x_1,\dots,x_{\check{n}})$ can be obtained from $p^{n_2}(\cdot|x_1,\dots,x_{\check{n}})$ through marginalisation; the latter essentially demands that these conditional probability mass functions should be connected with temporally consistent unconditional probability mass functions through Bayes's Rule.

A common assumption about the variables X_n is that they are *exchangeable*. De Finetti's famous Representation Theorem [11, 4] then states that the temporally consistent and coherent conditional and unconditional predictive probability mass functions associated with a countably infinite exchangeable sequence of variables in A are completely characterised by[1] a unique probability measure on the Borel sets of the simplex of all probability mass functions on A, called its *representation*.

This leads us to the central problem of predictive inference: since there is an infinity of such probability measures on the simplex, which one does a subject choose in a particular context, and how can a given choice be motivated and justified? The subjectivists of de Finetti's persuasion would answer that this question needs no answer: a subject's personal predictive probabilities are entirely his, and temporal consistency and coherence are the only requirements he should heed. Proponents of the logicist approach to predictive inference would try enunciating general inference principles in order to narrow down, and hopefully eliminate entirely, the possible choices for the representing probability measures on the simplex. Our point of view holds a compromise between the subjectivist and logicist positions: it should be possible for a subject to make assessments for certain predictive probabilities, and to combine these with certain inference principles he finds reasonable. Although this is not the topic of the present conference paper, the inference systems we introduce in Section 6 provide an elegant framework and tools for making conservative predictive inferences that combine (local) subjective probability assessments with (general) inference principles.

This idea of *conservative probabilistic inference* brings us to a central idea in de Finetti's approach to probability [13]: a subject should be able to make certain probability assessments, and we can then consider these as bounds on so-called precise probability models. Calculating such most conservative but tightest bounds is indeed what de Finetti's Fundamental Theorem of Prevision [13, 19] is about. The theory of imprecise probabilities [30, 25, 28] looks at conservative probabilistic inference precisely in this way: how can we calculate as efficiently as possible the consequences—in the sense of most conservative tightest bounds—of making certain probability assessments. One advantage of imprecise probability models is that they allow for *imprecision*, or in other words, the use of *partial* probability assessments using bounding *inequalities* rather than equalities. In Section 2, we

[1] … unless the observed sequence has probability zero.

give a concise overview of the relevant ideas, models and techniques in the field of imprecise probabilities.

The present paper, then, can be described as an application of ideas in imprecise probabilities to predictive inference. Its aim is to study—and develop a general framework for dealing with—coherent predictive inference using imprecise probability models. Using such models will also allow us to represent a subject's indecision, which we believe is a natural state to be in when knowing, or having learned little, about the problem at hand. It seems important to us that theories of learning under uncertainty in general, and predictive inference in particular, start out with conservative, very imprecise and indecisive models when little has been learned, and become more precise and decisive as more observations come in.

Our work here builds on, but manages to reach much further than, an earlier paper by one of the authors [7]. The main reason why it does so, is that we are now in a position to use a very powerful mathematical language to represent imprecise-probabilistic inferences: Walley's [28] coherent sets of desirable gambles. Here, the primitive notions are not probabilities of events, nor expectations of random variables. The focus is rather on the question whether a gamble, or a risky transaction, is desirable to a subject—strictly preferred to the zero transaction, or status quo. And a basic belief model is now not a probability measure or lower prevision, but a *set of desirable gambles*.

Let us briefly summarise why, in the present paper, we work with such sets as our basic uncertainty models for doing conservative probabilistic inference. Most importantly, and as we shall see in Sections 2 and 3, marginalisation and conditioning are especially straightforward, and there are no issues whatsoever with conditioning on sets of (lower) probability zero. Furthermore, sets of desirable gambles provide an extremely expressive and general framework: it encompasses and subsumes as special cases both classical (or 'precise') probabilistic inference and inference in classical propositional logic [6].

So, now that we have argued why we want to use sets of desirable gambles to extend the existing probabilistic theory of predictive inference, let us explain in some detail how we intend to go about doing this. The basic building blocks are introduced in Sections 2–8. As already indicated above, we give an overview of relevant notions and results concerning our imprecise probability model of choice—coherent sets of desirable gambles—in Section 2. In particular, we explain how to use them for conservative inference as well as conditioning; how to derive more commonly used models, such as lower previsions and lower probabilities, from them; and how they relate to precise probability models.

In Section 3, we explain how we can describe a subject's beliefs about a sequence of variables in terms of predictive sets of desirable gambles, and the derived notion of predictive lower previsions. These imprecise probability models generalise the above-mentioned predictive probability mass functions $p^{\hat{n}}(\cdot|x_1,\dots,x_{\check{n}})$, and they constitute the basic tools we shall be working with. We also explain what are the proper formulations for the above-mentioned temporal consistency and coherence requirements in this more general context.

In Section 4, we discuss a number of inference principles that we believe could be reasonably imposed on predictive inferences, and we show how to represent them mathematically in terms of predictive sets of desirable gambles and lower previsions. *Representation insensitivity* means that predictive inferences remain essentially unchanged when we transform the set of categories, or in other words that they are essentially insensitive to the choice of representation—the category set. Another inference principle we look at imposes the so-called *specificity* property: when predictive inference is specific, then for a specific question involving a restricted number of categories, a more general model can be replaced by a more specific model that deals only with the categories of interest, and will produce the same relevant inferences [2].

The next important step is taken in Section 5, where we recall from the literature [9, 8] how to deal with exchangeability when our predictive inference models are imprecise. We recall that de Finetti's Representation Theorem can be significantly generalised. In this case, the temporal consistent and coherent predictive sets of desirable gambles are completely characterised by a set of (multivariate) polynomials on the simplex of all probability mass functions on the category set. This set of polynomials must satisfy a number of properties, which taken together define the notion of *Bernstein coherence*. It serves completely the same purpose as the representing probability measure: it completely determines, and conveniently and densely summarises, all predictive inferences. This is the reason why the rest of the developments in the paper are expressed in terms of such Bernstein coherent sets of polynomials.

We introduce coherent inference systems in Section 6 as maps that associate with any finite set of categories a Bernstein coherent set of polynomials on the simplex of probability mass functions on that set. The inference principles in Section 4 impose connections between predictive inferences for different category sets, so we can represent such inference principles mathematically as restrictions on coherent inference systems, which is the main topic of Section 7.

The material in Sections 8–10 shows, by producing explicit examples, that there are quite a few different types—even uncountable infinities—of coherent inference systems that are both representation insensitive and specific. We discuss the vacuous inference system in Section 8, the family of IDMM inference systems in Section 9 and the Haldane inference system in Section 10.

In the Conclusion (Section 11) we point to a number of surprising consequences of our results, and discuss avenues for further research.

2 Imprecise probability models

In this section, we give a concise overview of imprecise probability models for representing, and making inferences and decisions under, uncertainty.

We shall focus on sets of desirable gambles as our uncertainty models of choice, because they are the most powerful, expressive and general models at hand, because they are very intuitive to work with—though unfortunately less familiar to most people not closely involved in the field—, and, very importantly, because they avoid problems with conditioning on sets of (lower) probability zero. For more details, we refer to Refs. [1, 10, 8, 21, 28]. We shall of course also briefly mention derived results in terms of the more familiar language of (lower) previsions and probabilities.

We consider a variable X that assumes values in some possibility space A. We model a subject's beliefs about the value of X by looking at which gambles on this variable the subject finds *desirable*, meaning that he strictly prefers them to the zero gamble—the status quo. This is a very general approach, that extends the usual rationalist and subjectivist approach to probabilistic modelling to allow for indecision and imprecision.

A *gamble* is a (bounded) real-valued function f on A. It is interpreted as an uncertain reward $f(X)$ that depends on the value of X, and is expressed in units of some predetermined linear utility. It represents the reward the subject gets in a transaction where first the actual value x of X is determined, and then the subject receives the amount of utility $f(x)$—which may be negative, meaning he has to pay it. Throughout the paper, we shall use the device of writing $f(X)$ when we want to make clear what variable the gamble f depends on. *Events* are subsets of the possibility space A. With any event $B \subseteq A$ we can associate a special gamble $\mathbb{I}_B$, called its *indicator*, which assumes the value 1 on B and 0 elsewhere.

We denote the set of all gambles on A by $\mathscr{G}(A)$. It is a linear space under point-wise addition of gambles, and point-wise multiplication of gambles with real numbers. For any subset $\mathscr{A}$ of $\mathscr{G}(A)$, $\operatorname{posi}(\mathscr{A})$ is the set of all positive linear combinations of gambles in $\mathscr{A}$: $\operatorname{posi}(\mathscr{A}) := \{\sum_{k=1}^{n} \lambda_k f_k \colon f_k \in \mathscr{A}, \lambda_k \in \mathbb{R}_{>0}, n \in \mathbb{N}\}$. Here, $\mathbb{N}$ is the set of natural numbers (without zero), and $\mathbb{R}_{>0}$ is the set of all positive real numbers. A *convex cone* of gambles is a subset $\mathscr{A}$ of $\mathscr{G}(A)$ that is closed under positive linear combinations, meaning that $\operatorname{posi}(\mathscr{A}) = \mathscr{A}$. For any two gambles f and g on A, we write '$f \geq g$' if $(\forall x \in A) f(x) \geq g(x)$, and '$f > g$' if $f \geq g$ and $f \neq g$. A gamble $f > 0$ is called *positive*. A gamble $g \leq 0$ is called *non-positive*. $\mathscr{G}_{>0}(A)$ denotes the convex cone of all positive gambles, and $\mathscr{G}_{\leq 0}(A)$ the convex cone of all non-positive gambles.

We collect the gambles that a subject finds desirable—strictly prefers to the zero gamble—into his *set of desirable gambles*, and we shall take such sets as our basic uncertainty models. Of course, they have to satisfy certain rationality criteria:

Definition 1 (Coherence). A set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(A)$ is called *coherent* if it satisfies the following requirements:

D1. $0 \notin \mathscr{D}$;
D2. $\mathscr{G}_{>0}(A) \subseteq \mathscr{D}$;
D3. $\mathscr{D} = \operatorname{posi}(\mathscr{D})$.

Requirement D3 turns $\mathscr{D}$ into a *convex cone*. Due to D2, it includes $\mathscr{G}_{>0}(A)$; by D1–D3, it *avoids non-positivity*:

D4. if $f \leq 0$ then $f \notin \operatorname{posi}(\mathscr{D})$, or equivalently $\mathscr{G}_{\leq 0}(A) \cap \operatorname{posi}(\mathscr{D}) = \emptyset$.

$\mathscr{G}_{>0}(A)$ is the smallest coherent subset of $\mathscr{G}(A)$. This so-called *vacuous model* therefore reflects minimal commitments on the part of the subject: if he knows absolutely nothing about the likelihood of the different outcomes, he will only strictly prefer to zero those gambles that never decrease his wealth and have some possibility of increasing it.

Let us suppose that our subject has a coherent set $\mathscr{D}$ of desirable gambles on A, expressing his beliefs about the value that a variable X assumes in A. We can then ask what his so-called *updated* set $\mathscr{D}\rfloor B$ of desirable gambles on B would be were he to receive the additional information—and nothing more—that X actually belongs to some subset B of A. The *updating*, or *conditioning*, *rule* for sets of desirable gambles states that:

$$g \in \mathscr{D}\rfloor B \Leftrightarrow g\mathbb{I}_B \in \mathscr{D} \text{ for all gambles } g \text{ on } B. \tag{1}$$

It states that the gamble g is desirable to a subject were he to observe that $X \in B$ if and only if the *called-off gamble* $g\mathbb{I}_B$ is desirable to him. This called-off gamble $g\mathbb{I}_B$ is the gamble on the variable X that gives a zero reward—is called off—unless $X \in B$, and in that case reduces to the gamble g on the new possibility space B. The updated set $\mathscr{D}\rfloor B$ is a set of desirable gambles on B that is still coherent, provided that $\mathscr{D}$ is [8]. We refer to Refs. [21, 10, 22] for detailed discussions of updating sets of desirable gambles.

We now use coherent sets of desirable gambles to introduce derived concepts, such as coherent lower previsions, and probabilities. Given a coherent set of desirable gambles $\mathscr{D}$, the functional $\underline{P}$ defined on $\mathscr{G}(A)$ by

$$\underline{P}(f) := \sup\{\mu \in \mathbb{R} \colon f - \mu \in \mathscr{D}\} \text{ for all } f \in \mathscr{G}(A), \tag{2}$$

is a *coherent lower prevision* [25, Theorem 3.8.1]. The conjugate upper prevision $\overline{P}$ is defined by $\overline{P}(f) := \inf\{\mu \in \mathbb{R} \colon \mu - f \in \mathscr{D}\} = -\underline{P}(-f)$. For any gamble f, $\underline{P}(f)$ is called the *lower prevision* of f, and for any event B, $\underline{P}(\mathbb{I}_B)$ is also denoted by $\underline{P}(B)$, and called the *lower probability* of B. Similarly for upper previsions and upper probabilities.

The coherent conditional model $\mathscr{D}\rfloor B$, with B a non-empty subset of A, induces a *conditional lower prevision* $\underline{P}(\cdot|B)$ on $\mathscr{G}(B)$, by applying Equation (2):

$$\underline{P}(g|B) := \sup\{\mu \in \mathbb{R} \colon g - \mu \in \mathscr{D}\rfloor B\} = \sup\{\mu \in \mathbb{R} \colon [g - \mu]\mathbb{I}_B \in \mathscr{D}\} \quad \text{for all gambles } g \text{ on } B. \tag{3}$$

It is not difficult to show [25] that $\underline{P}$ and $\underline{P}(\cdot|B)$ are related through the following coherence condition:

$$\underline{P}([g - \underline{P}(g|B)]\mathbb{I}_B) = 0 \text{ for all } g \in \mathscr{G}(B), \tag{GBR}$$

called the *Generalised Bayes Rule*. This rule allows us to infer $\underline{P}(\cdot|B)$ uniquely from $\underline{P}$, provided that $\underline{P}(B) > 0$. Otherwise, there are an infinity of coherent lower previsions $\underline{P}(\cdot|B)$ that are coherent with $\underline{P}$ in the sense that they satisfy (GBR).

Coherent sets of desirable gambles are more informative than coherent lower previsions: a gamble with positive lower prevision is always desirable and one with a negative lower prevision never, but a gamble with zero lower prevision lies on the border of the set of desirable gambles, and the lower prevision does not generally provide information about the desirability of such gambles. If such border behaviour is important—and it is when dealing with conditioning on events with zero (lower) probability [28, 21, 22, 10]—it is useful to work with sets of desirable gambles rather than lower previsions, because as Equations (1) and (3) tell us, they allow us to derive unique conditional models from unconditional ones.

When the lower and the upper prevision coincide on all gambles, then the real functional P defined on $\mathscr{G}(A)$ by $P(f) := \underline{P}(f) = \overline{P}(f)$ for all $f \in \mathscr{G}(A)$ is a *linear prevision*. In the particular case that A is finite, this means that it corresponds to the expectation operator associated with a probability mass function p: $P(f) = \sum_{x \in A} f(x)p(x) =: E_p(f)$, where $p(x) := P(\mathbb{I}_{\{x\}})$ for all $x \in A$.

3 Predictive inference

Predictive inference, in the specific sense we are focussing on here, considers a number of variables $X_1, \dots, X_n$ assuming values in the same category set A—we define a *category set* as any non-empty *finite* set. We start our discussion of predictive inference models in the most general and representationally powerful language: coherent sets of desirable gambles, as introduced in the previous section.

Predictive inference assumes generally that a number $\check{n}$ of observations have been made, so we know the values $\check{\boldsymbol{x}} = (x_1, \dots, x_{\check{n}})$ of the first $\check{n}$ variables $X_1, \dots, X_{\check{n}}$. Based on this *observation sample* $\check{\boldsymbol{x}}$, a subject then has a posterior *predictive model* $\mathscr{D}_A^{\hat{n}} \rfloor \check{\boldsymbol{x}}$ for the values that the next $\hat{n}$ variables $X_{\check{n}+1}, \dots, X_{\check{n}+\hat{n}}$ assume in $A^{\hat{n}}$. $\mathscr{D}_A^{\hat{n}} \rfloor \check{\boldsymbol{x}}$ is a coherent set of desirable gambles $f(X_{\check{n}+1}, \dots, X_{\check{n}+\hat{n}})$ on $A^{\hat{n}}$. Here we assume that $\hat{n} \in \mathbb{N}$. On the other hand, we want to allow that $\check{n} \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, which is the set of all natural numbers with zero: we also want to be able to deal with the case where no previous observations have been made. In that case, we call the corresponding model $\mathscr{D}_A^{\hat{n}}$ a *prior predictive model*. Of course, technically speaking, $\check{n} + \hat{n} \leq n$.

As we said, the subject may also have a prior, unconditional model, for when no observations have yet been made. In its most general form, this will be a coherent set $\mathscr{D}_A^n$ of desirable gambles $f(X_1, \dots, X_n)$ on A^n, for some $n \in \mathbb{N}$. Our subject may also have a coherent set $\mathscr{D}_A^{\hat{n}}$ of desirable gambles $f(X_1, \dots, X_{\hat{n}})$ on $A^{\hat{n}}$, where $\hat{n} \leq n$; and the sets $\mathscr{D}_A^{\hat{n}}$ and $\mathscr{D}_A^n$ must then be related to each other through the following *marginalisation*, or *temporal consistency*, requirement:

$$f(X_1, \dots, X_{\hat{n}}) \in \mathscr{D}_A^{\hat{n}} \Leftrightarrow f(X_1, \dots, X_{\hat{n}}) \in \mathscr{D}_A^n \text{ for all gambles } f \text{ on } A^{\hat{n}}. \tag{4}$$

In this expression, and throughout this paper, we identify a gamble f on $A^{\hat{n}}$ with its *cylindrical extension* f' on A^n, defined by $f'(x_1, \dots, x_{\hat{n}}, \dots, x_n) := f(x_1, \dots, x_{\hat{n}})$ for all $(x_1, \dots, x_n) \in A^n$. If we introduce the marginalisation operator $\operatorname{marg}_{\hat{n}}(\cdot) :=$

$\cdot \cap \mathscr{G}(A^k)$, then the temporal consistency condition can also be rewritten simply as $\mathscr{D}^{\hat{n}}_A = \operatorname{marg}_{\hat{n}}(\mathscr{D}^n_A) = \mathscr{D}^n_A \cap \mathscr{G}(A^{\hat{n}})$.

Prior (unconditional) predictive models $\mathscr{D}^n_A$ and posterior (conditional) ones $\mathscr{D}^{\hat{n}}_A \rfloor \check{\boldsymbol{x}}$ must also be related through the following *updating* requirement:

$$f(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}}) \in \mathscr{D}^{\hat{n}}_A \rfloor \check{\boldsymbol{x}} \Leftrightarrow f(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}})\mathbb{I}_{\{\check{\boldsymbol{x}}\}}(X_1,\dots,X_{\check{n}}) \in \mathscr{D}^n_A \\ \text{for all gambles } f \text{ on } A^{\hat{n}}, \quad (5)$$

which is a special case of Equation (1): the gamble $f(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}})$ is desirable after observing a sample $\check{\boldsymbol{x}}$ if and only if the gamble $f(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}})\mathbb{I}_{\{\check{\boldsymbol{x}}\}}(X_1,\dots,X_{\check{n}})$ is desirable before any observations are made. This called-off gamble is the gamble that gives zero reward—is called off—unless the first $\check{n}$ observations are $\check{\boldsymbol{x}}$, and in that case reduces to the gamble $f(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}})$ on the variables $X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}}$. The updating requirement is a generalisation of Bayes's Rule for updating, and in fact reduces to it when the sets of desirable gambles lead to (precise) probability mass functions [28, 6]. But contrary to Bayes's Rule for probability mass functions, the updating rule (5) for coherent sets of desirable gambles clearly does not suffer from problems when the conditioning event has (lower) probability zero: it allows us to infer a unique conditional model from an unconditional one, regardless of the (lower or upper) probability of the conditioning event.

As explained in Section 2, we can use the relationship (2) to derive *prior* (unconditional) *predictive lower previsions* $\underline{P}^{\hat{n}}_A(\cdot)$ on $\mathscr{G}(A^{\hat{n}})$ from the prior sets $\mathscr{D}^{\hat{n}}_A$ through:

$$\underline{P}^{\hat{n}}_A(f) := \sup\left\{\mu \in \mathbb{R} \colon f - \mu \in \mathscr{D}^{\hat{n}}_A\right\} \text{ for all gambles } f \text{ on } A^{\hat{n}},$$

and *posterior* (conditional) *predictive lower previsions* $\underline{P}^{\hat{n}}_A(\cdot|\check{\boldsymbol{x}})$ on $\mathscr{G}(A^{\hat{n}})$ from the posterior sets $\mathscr{D}^{\hat{n}}_A \rfloor \check{\boldsymbol{x}}$ through:

$$\underline{P}^{\hat{n}}_A(f|\check{\boldsymbol{x}}) := \sup\left\{\mu \in \mathbb{R} \colon f - \mu \in \mathscr{D}^{\hat{n}}_A \rfloor \check{\boldsymbol{x}}\right\} \text{ for all gambles } f \text{ on } A^{\hat{n}}.$$

We also want to condition predictive lower previsions on the additional information that $(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}}) \in B^{\hat{n}}$, where B is some proper subset of A. Using the ideas in Sections 2, this leads for instance to the following lower prevision:

$$\underline{P}^{\hat{n}}_A(g|\check{\boldsymbol{x}},B^{\hat{n}}) := \sup\left\{\mu \in \mathbb{R} \colon [g-\mu]\mathbb{I}_{B^{\check{n}}} \in \mathscr{D}^{\hat{n}}_A \rfloor \check{\boldsymbol{x}}\right\} \text{ for all gambles } g \text{ on } B^{\hat{n}}, \quad (6)$$

which is the lower prevision $\underline{P}^{\hat{n}}_A(\cdot|\check{\boldsymbol{x}})$ conditioned on the event $B^{\hat{n}}$.

4 Principles for predictive inference

So far, we have introduced coherence, marginalisation and updating as basic requirements of rationality that prior and posterior predictive inference models must satisfy. In addition to these, we now also consider a number of further conditions, which have been suggested by a number of authors as reasonable properties—or requirements—for predictive inference models.

We shall call *representation insensitivity* the combination of pooling, renaming and category permutation invariance; see Ref. [7] for more information. It means that predictive inferences remain essentially unchanged when we transform the set of categories, or in other words that they are essentially insensitive to the choice of representation—the category set. It is not difficult to see that representation insensitivity can be formally characterised as follows. Consider two category sets A and B such that there is a so-called *relabelling map* $\rho\colon A\to B$ that is *onto*, i.e. such that $B=\rho(A):=\{\rho(x)\colon x\in A\}$. Then with a sample $\boldsymbol{x}$ in A^n, there corresponds a transformed sample $\rho\boldsymbol{x}:=(\rho(x_1),\dots,\rho(x_n))$ in B^n. And with any gamble f on B^n there corresponds a gamble $f\circ\rho$ on A^n.

Representation insensitivity: For all category sets A and B such that there is an onto map $\rho\colon A\to B$, all $\check{n},\hat{n}\in\mathbb{N}$ considered, all $\check{\boldsymbol{x}}\in A^{\check{n}}$ and all gambles f on $B^{\hat{n}}$:

$$\underline{P}_A^{\hat{n}}(f\circ\rho)=\underline{P}_B^{\hat{n}}(f)\text{ and }\underline{P}_A^{\hat{n}}(f\circ\rho|\check{\boldsymbol{x}})=\underline{P}_B^{\hat{n}}(f|\rho\check{\boldsymbol{x}}),\tag{RI1}$$

or alternatively, and more generally, in terms of predictive sets of desirable gambles:

$$f\circ\rho\in\mathscr{D}_A^{\hat{n}}\Leftrightarrow f\in\mathscr{D}_B^{\hat{n}}\text{ and }f\circ\rho\in\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{x}}\Leftrightarrow f\in\mathscr{D}_B^{\hat{n}}\rfloor\rho\check{\boldsymbol{x}}.\tag{RI2}$$

There is another peculiar, but in our view intuitively appealing, potential property of predictive inferences. Assume that in addition to observing a sample of observations $\check{\boldsymbol{x}}$ of $\check{n}$ observations in a category set A, our subject comes to know or determine in some way that the $\hat{n}$ following observations will belong to a proper subset B of A, and nothing else—we might suppose for instance that an observation of $(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}})$ has been made, but that it is imperfect, and only allows him to conclude that $(X_{\check{n}+1},\dots,X_{\check{n}+\hat{n}})\in B^{\hat{n}}$.

We can then make the following requirement, which uses models conditioned on the event $B^{\hat{n}}$, as introduced through Equations (1), (3) and (6).

Specificity: For all category sets A and B such that $B\subseteq A$, all $\check{n},\hat{n}\in\mathbb{N}$ considered, all $\check{\boldsymbol{x}}\in A^{\check{n}}$ and all gambles f on $B^{\hat{n}}$:

$$\underline{P}_A^{\hat{n}}(f|B^{\hat{n}})=\underline{P}_B^{\hat{n}}(f)\text{ and }\underline{P}_A^{\hat{n}}(f|\check{\boldsymbol{x}},B^{\hat{n}})=\underline{P}_B^{\hat{n}}(f|\check{\boldsymbol{x}}{\downarrow_B}),\tag{SP1}$$

or alternatively, and more generally, in terms of predictive sets of desirable gambles:

$$f\mathbb{I}_{B^{\hat{n}}}\in\mathscr{D}_A^{\hat{n}}\Leftrightarrow f\in\mathscr{D}_B^{\hat{n}}\text{ and }f\mathbb{I}_{B^{\hat{n}}}\in\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{x}}\Leftrightarrow f\in\mathscr{D}_B^{\hat{n}}\rfloor\check{\boldsymbol{x}}{\downarrow_B},\tag{SP2}$$

where $\check{\boldsymbol{x}}{\downarrow_B}$ is the tuple of observations obtained by eliminating from the tuple $\check{\boldsymbol{x}}$ all observations not in B. In these expressions, when $\check{\boldsymbol{x}}{\downarrow_B}$ is the empty tuple, so when no observations in $\check{\boldsymbol{x}}$ are in B, the 'posterior' predictive model is simply taken to reduce to the 'prior' predictive model. Specificity [2, 3, 24] means that *the predictive inferences that a subject makes are the same as the ones he would get by focussing on the category set B, and at the same time discarding all the* previous *observations* *producing values outside B, in effect only retaining the observations that were inside B!* It is as if knowing that the future observations belong to B allows our subject to ignore all the previous observations that happened to lie outside B.

5 Adding exchangeability to the picture

We are now, for the remainder of this paper, going to add two additional assumptions. The *first assumption* is that we are dealing with a *countably infinite sequence* of variables $X_1, \dots, X_n, \dots$ that assume values in the same category set A. For our predictive inference models, this means that there is a sequence $\mathscr{D}_A^n$ of coherent sets of desirable gambles on A^n, $n \in \mathbb{N}$. The *second assumption* is that this sequence of variables is *exchangeable*, which means, roughly speaking, that the subject believes that the order in which these variables are observed, or present themselves, has no influence on the decisions and inferences he will make regarding these variables.

In this section, we explain succinctly how to deal with these assumptions technically, and what their consequences are for the predictive models we are interested in. For a detailed discussion and derivation of the results presented here, we refer to Refs. [9, 8].

We begin with some useful notation, which will be employed numerous times in what follows. Consider any element $\boldsymbol{\alpha} \in \mathbb{R}^A$. We consider $\boldsymbol{\alpha}$ as an *A-tuple*, with as many (real) components $\alpha_x \in \mathbb{R}$ as there are categories x in A. For any subset $B \subseteq A$, we then denote by $\alpha_B := \sum_{x \in B} \alpha_x$ the sum of its components over B.

Consider an arbitrary $n \in \mathbb{N}$. We denote by $\boldsymbol{x} = (x_1, \dots, x_n)$ a generic, arbitrary element of A^n. $\mathscr{P}^n$ is the set of all permutations π of the index set $\{1, \dots, n\}$. With any such permutation π, we can associate a permutation of A^n, also denoted by π, and defined by $(\pi\boldsymbol{x})_k := x_{\pi(k)}$, or in other words, $\pi(x_1, \dots, x_n) := (x_{\pi(1)}, \dots, x_{\pi(n)})$. Similarly, we lift π to a permutation π^t of $\mathscr{G}(A^n)$ by letting $\pi^t f := f \circ \pi$, so $(\pi^t f)(\boldsymbol{x}) := f(\pi\boldsymbol{x})$. The permutation invariant atoms $[\boldsymbol{x}] := \{\pi\boldsymbol{x} \colon \pi \in \mathscr{P}^n\}$, $\boldsymbol{x} \in A^n$ are the smallest permutation invariant subsets of A^n.

We now introduce the *counting map* $\boldsymbol{T} \colon A^n \to \mathscr{N}_A^n \colon \boldsymbol{x} \mapsto \boldsymbol{T}(\boldsymbol{x})$, where the *count vector* $\boldsymbol{T}(\boldsymbol{x})$ is the A-tuple with components $T_z(\boldsymbol{x}) := |\{k \in \{1, \dots, n\} \colon x_k = z\}|$ for all $z \in A$, and the set of possible *count vectors* for n observations in A is given by $\mathscr{N}_A^n := \left\{\boldsymbol{m} \in \mathbb{N}_0^A \colon m_A = n\right\}$. So $T_z(\boldsymbol{x})$ is the number of times the category z appears in the sample $\boldsymbol{x}$. If $\boldsymbol{m} = \boldsymbol{T}(\boldsymbol{x})$, then $[\boldsymbol{x}] = \{\boldsymbol{y} \in A^n \colon \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{m}\}$, so the atom $[\boldsymbol{x}]$ is completely determined by the single count vector $\boldsymbol{m}$ of all its elements, and is therefore also denoted by $[\boldsymbol{m}]$.

We also consider the linear expectation operator $\mathrm{Hy}_A^n(\cdot|\boldsymbol{m})$ associated with the uniform distribution on the invariant atom $[\boldsymbol{m}]$:

$$\mathrm{Hy}_A^n(f|\boldsymbol{m}) := \frac{1}{|[\boldsymbol{m}]|} \sum_{\boldsymbol{x} \in [\boldsymbol{m}]} f(\boldsymbol{x}) \text{ for all gambles } f \text{ on } A^n,$$

where the number of elements $\nu(\boldsymbol{m}) := |[\boldsymbol{m}]|$ in the invariant atom $[\boldsymbol{m}]$ is given by the *multinomial coefficient*:

$$\nu(\boldsymbol{m}) = \binom{m_A}{\boldsymbol{m}} = \binom{n}{\boldsymbol{m}} := \frac{n!}{\prod_{z\in A} m_z!}.$$

This expectation operator characterises a (multivariate) *hyper-geometric distribution* [17, Section 39.2], associated with random sampling without replacement from an urn with n balls of types $z \in A$, whose composition is characterised by the count vector $\boldsymbol{m}$. This hyper-geometric expectation operator can also be seen as a linear transformation Hy_A^n between the linear space $\mathscr{G}(A^n)$ and the generally much lower-dimensional linear space $\mathscr{G}(\mathscr{N}_A^n)$, turning a gamble f on A^n into a so-called *count gamble* $\mathrm{Hy}_A^n(f) := \mathrm{Hy}_A^n(f|\cdot)$ on count vectors.

Next, we consider the simplex Σ_A of all probability mass functions $\boldsymbol{\theta}$ on A: $\Sigma_A := \left\{\boldsymbol{\theta} \in \mathbb{R}^A \colon \boldsymbol{\theta} \geq 0 \text{ and } \theta_A = 1\right\}$. With a probability mass function $\boldsymbol{\theta} \in \Sigma_A$ on A, there corresponds the following *multinomial expectation* operator $\mathrm{Mn}_A^n(\cdot|\boldsymbol{\theta})$:[2]

$$\mathrm{Mn}_A^n(f|\boldsymbol{\theta}) := \sum_{\boldsymbol{x}\in A^n} f(\boldsymbol{x}) \prod_{z\in A} \theta_z^{T_z(\boldsymbol{x})} \text{ for all gambles } f \text{ on } A^n,$$

which characterises the multinomial distribution, associated with n independent trials of an experiment with possible outcomes in A and probability mass function $\boldsymbol{\theta}$. Observe that $\mathrm{Mn}_A^n(f|\boldsymbol{\theta}) = \sum_{\boldsymbol{m}\in\mathscr{N}_A^n} \mathrm{Hy}_A^n(f|\boldsymbol{m})\nu(\boldsymbol{m})\prod_{z\in A}\theta_z^{m_z} = \mathrm{CoMn}_A^n(\mathrm{Hy}_A^n(f)|\boldsymbol{\theta})$, where we used the so-called *count multinomial expectation* operator:

$$\mathrm{CoMn}_A^n(g|\boldsymbol{\theta}) := \sum_{\boldsymbol{m}\in\mathscr{N}_A^n} g(\boldsymbol{m})\nu(\boldsymbol{m}) \prod_{z\in A} \theta_z^{m_z} \text{ for all gambles } g \text{ on } \mathscr{N}_A^n. \tag{7}$$

Let us introduce the notation $\mathscr{N}_A := \bigcup_{m\in\mathbb{N}} \mathscr{N}_A^m$ for the set of all possible count vectors, corresponding to samples of at least one observation. $\mathscr{N}_A^0$ is the singleton containing only the null count vector $\mathbf{0}$, all of whose components are zero. Then $\bigcup_{m\in\mathbb{N}_0} \mathscr{N}_A^m = \mathscr{N}_A \cup \{\mathbf{0}\}$ is the set of all possible count vectors. For any such count vector $\boldsymbol{m} \in \mathscr{N}_A \cup \{\mathbf{0}\}$, we consider the (multivariate) *Bernstein basis polynomial* $\mathrm{B}_{A,\boldsymbol{m}}$ of degree m_A on Σ_A, defined by:

$$\mathrm{B}_{A,\boldsymbol{m}}(\boldsymbol{\theta}) := \nu(\boldsymbol{m}) \prod_{z\in A} \theta_z^{m_z} = \binom{m_A}{\boldsymbol{m}} \prod_{z\in A} \theta_z^{m_z} \text{ for all } \boldsymbol{\theta} \in \Sigma_A. \tag{8}$$

In particular, of course, $\mathrm{B}_{A,\mathbf{0}} = 1$. Any linear combination p of Bernstein basis polynomials of degree $n \geq 0$ is a (multivariate) polynomial (gamble) on Σ_A, whose degree $\deg(p)$ is at most n.[3] We denote the linear space of all these polynomials

[2] To avoid confusion, we make a (perhaps non-standard) distinction between the multinomial expectation, which is associated with sequences of observations, and the count multinomial expectation, associated with their count vectors.

[3] The degree may be smaller than n because the sum of all Bernstein basis polynomials of fixed degree is one. Strictly speaking, these polynomials p are restrictions to Σ_A of multivariate polynomials

of degree up to n by $\mathscr{V}^n(A)$. Of course, polynomials of degree zero are simply real constants. For any $n \geq 0$, we can introduce a linear isomorphism CoMn_A^n between the linear spaces $\mathscr{G}(\mathscr{N}_A^n)$ and $\mathscr{V}^n(A)$: with any gamble g on $\mathscr{N}_A^n$, there corresponds a polynomial $\mathrm{CoMn}_A^n(g) := \mathrm{CoMn}_A^n(g|\cdot) = \sum_{\boldsymbol{m}\in\mathscr{N}_A^n} g(\boldsymbol{m})\mathrm{B}_{A,\boldsymbol{m}}$ in $\mathscr{V}^n(A)$, and conversely, for any polynomial $p \in \mathscr{V}^n(A)$ there is a unique gamble b_p^n on $\mathscr{N}_A^n$ such that $p = \mathrm{CoMn}_A^n(b_p^n)$ [8].[4] We denote by $\mathscr{V}(A) := \bigcup_{n\in\mathbb{N}_0} \mathscr{V}^n(A)$ the linear space of all (multivariate) polynomials on Σ_A, of arbitrary degree.

A set $\mathscr{H}_A \subseteq \mathscr{V}(A)$ of polynomials on Σ_A is called *Bernstein coherent* if it satisfies the following properties:

B1. $0 \notin \mathscr{H}_A$;
B2. $\mathscr{V}^+(A) \subseteq \mathscr{H}_A$;
B3. $\mathrm{posi}(\mathscr{H}_A) = \mathscr{H}_A$.

Here, $\mathscr{V}^+(A)$ is the set of *Bernstein positive* polynomials on Σ_A: those polynomials p such that $p(\boldsymbol{\theta}) > 0$ for all $\boldsymbol{\theta}$ in the interior $\mathrm{int}(\Sigma_A) := \{\boldsymbol{\theta} \in \Sigma_A \colon (\forall x \in A)\theta_x > 0\}$ of Σ_A. As a consequence, for the set $\mathscr{V}_0^-(A) := -\mathscr{V}^+(A) \cup \{0\}$ of *Bernstein non-positive* polynomials:

B4. $\mathscr{V}_0^-(A) \cap \mathscr{H}_A = \emptyset$.

We are now ready to deal with exchangeability. We shall give a definition for coherent sets of desirable gambles that generalises de Finetti's definition [11, 13], and which allows for a generalisation of his Representation Theorem.

First of all, fix $n \in \mathbb{N}$. Then the subject considers the variables $X_1,\dots,X_n$ to be exchangeable when he does not distinguish between any gamble f on A^n and its permuted version $\pi^t f$, or in other words, if the gamble $f - \pi^t f$ is equivalent to the zero gamble for—or *indifferent* to—him. This means that he has a so-called *set of indifferent gambles*: $\mathscr{I}_A^n := \{f - \pi^t f \colon f \in \mathscr{G}(A^n) \text{ and } \pi \in \mathscr{P}^n\}$. If the subject also has a coherent set of desirable gambles $\mathscr{D}_A^n$, then this set must be compatible with the set of indifferent gambles $\mathscr{I}_A^n$, in the sense that it must satisfy the rationality requirement $\mathscr{D}_A^n + \mathscr{I}_A^n = \mathscr{D}_A^n$ [8, 23]. We then say that the sequence $X_1,\dots,X_n$, and the model $\mathscr{D}_A^n$, are *exchangeable*. Next, the countably infinite sequence of variables $X_1,\dots,X_n\dots$ is called exchangeable if all the finite subsequences $X_1,\dots,X_n$ are, for $n \in \mathbb{N}$. This means that all models $\mathscr{D}_A^n$, $n \in \mathbb{N}$ are exchangeable. They should of course also be temporally consistent.

Theorem 1 (Representation Theorem [8]). *The sequence of sets $\mathscr{D}_A^n$ of desirable gambles on A^n, $n \in \mathbb{N}$ is coherent, temporally consistent and exchangeable if and only if there is a Bernstein coherent set $\mathscr{H}_A$ of polynomials on Σ_A such that for all $\hat{n} \in \mathbb{N}$, all gambles f on $A^{\hat{n}}$, all $\check{\boldsymbol{m}} \in \mathscr{N}_A$ and all $\check{\boldsymbol{x}} \in [\check{\boldsymbol{m}}]$:*

$$f \in \mathscr{D}_A^{\hat{n}} \Leftrightarrow \mathrm{Mn}_A^{\hat{n}}(f) \in \mathscr{H}_A \text{ and } f \in \mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{x}} \Leftrightarrow \mathrm{Mn}_A^{\hat{n}}(f)\mathrm{B}_{A,\check{\boldsymbol{m}}} \in \mathscr{H}_A. \tag{9}$$

q on $\mathbb{R}^A$, called *representations* of p. For any p, there are multiple representations, with possibly different degrees. The smallest such degree is then called the degree $\deg(p)$ of p.

[4] Strictly speaking, Equation (7) only defines the count multinomial expectation operator CoMn_A^n for $n > 0$, but it is clear that the definition extends trivially to the case $n = 0$.

In that case this representation $\mathscr{H}_A$ *is unique and given by* $\mathscr{H}_A := \bigcup_{n\in\mathbb{N}} \mathrm{Mn}_A^n(\mathscr{D}_A^n)$.

The representation $\mathscr{H}_A$ is a set of polynomials that plays the same role as a density, or distribution function, on Σ_A in the precise-probabilistic case. It follows from Equation (9) that $\mathscr{H}_A$ *completely determines* all predictive inferences about the sequence of variables $X_1,\dots,X_n,\dots$, as it fixes all prior predictive models $\mathscr{D}_A^{\hat{n}}$ and all posterior predictive models $\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{x}}$.

Equation (9) also tells us that the posterior predictive models $\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{x}}$ only depend on the observed sequence $\check{\boldsymbol{x}}$ through the count vector $\check{\boldsymbol{m}} = \boldsymbol{T}(\check{\boldsymbol{x}})$: count vectors are *sufficient statistics* under exchangeability. For this reason, we shall from now on denote these posterior predictive models by $\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{m}}$ as well as by $\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{x}}$. Also, every now and then, we shall use $\mathscr{D}_A^{\hat{n}}\rfloor\boldsymbol{0}$ as an alternative notation for $\mathscr{D}_A^{\hat{n}}$.

An immediate but interesting consequence of Theorem 1 is that updating on observations preserves exchangeability: after observing the values of the first $\check{n}$ variables, with count vector $\check{\boldsymbol{m}}$, the remaining sequence of variables $X_{\check{n}+1}, X_{\check{n}+2},\dots$ is still exchangeable, and Equation 9 tells us that its representation is given by the Bernstein coherent set of polynomials $\mathscr{H}_A\rfloor\check{\boldsymbol{m}}$ defined by:

$$\mathscr{H}_A\rfloor\check{\boldsymbol{m}} := \left\{p \in \mathscr{V}(A)\colon \mathrm{B}_{A,\check{\boldsymbol{m}}}\, p \in \mathscr{H}_A\right\}. \tag{10}$$

For the special case $\check{\boldsymbol{m}} = \boldsymbol{0}$, we find that $\mathscr{H}_A\rfloor\boldsymbol{0} = \mathscr{H}_A$. Clearly, $\mathscr{H}_A\rfloor\check{\boldsymbol{m}}$ is completely determined by $\mathscr{H}_A$. One can consider $\mathscr{H}_A$ as a prior model on the parameter space Σ_A, and $\mathscr{H}_A\rfloor\check{\boldsymbol{m}}$ plays the role of the posterior that is derived from it. We see from Equations (9) and (10) that—similar to what happens in a precise-probabilistic setting—the multinomial distribution serves as a direct link between on the one hand, the 'prior' $\mathscr{H}_A$ and its prior predictive inference models $\mathscr{D}_A^{\hat{n}}$ and, on the other hand, the 'posterior' $\mathscr{H}_A\rfloor\check{\boldsymbol{m}}$ and its posterior predictive inference models $\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{m}}$. Recalling our convention for $\check{\boldsymbol{m}} = \boldsymbol{0}$, we can summarise this as follows: for all $\hat{n} \in \mathbb{N}$ and all $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\boldsymbol{0}\}$:

$$\mathscr{D}_A^{\hat{n}}\rfloor\check{\boldsymbol{m}} = \left\{f \in \mathscr{G}(A^{\hat{n}})\colon\ \mathrm{Mn}_A^{\hat{n}}(f) \in \mathscr{H}_A\rfloor\check{\boldsymbol{m}}\right\} \tag{11}$$

and, as an immediate consequence,

$$\underline{P}_A^{\hat{n}}(f|\check{\boldsymbol{m}}) = \sup\left\{\mu \in \mathbb{R}\colon\ \mathrm{Mn}_A^{\hat{n}}(f) - \mu \in \mathscr{H}_A\rfloor\check{\boldsymbol{m}}\right\} \text{ for all } f \in \mathscr{G}(A^{\hat{n}}). \tag{12}$$

The sets of desirable polynomials $\mathscr{H}_A$ are the fundamental models, as they allow us to determine the $\mathscr{H}_A\rfloor\check{\boldsymbol{m}}$ and all predictive models uniquely.

6 Inference systems

We have seen in the previous section that, once we fix a category set A, predictive inferences about exchangeable sequences assuming values in A are completely determined by a Bernstein coherent set $\mathscr{H}_A$ of polynomials on Σ_A. So if we had some way of associating a Bernstein coherent set $\mathscr{H}_A$ with every possible set of categories A, this would completely fix all predictive inferences. This leads us to the following definition.

Definition 2 (Inference systems). We denote by $\mathbb{F}$ the collection of all category sets, i.e. finite non-empty sets. An *inference system* is a map Φ that maps any category set $A \in \mathbb{F}$ to some set of polynomials $\Phi(A) = \mathscr{H}_A$ on Σ_A. An inference system Φ is *coherent* if for all category sets $A \in \mathbb{F}$, $\Phi(A)$ is a Bernstein coherent set of polynomials on Σ_A.

So, a coherent inference system is a way to systematically associate coherent predictive inferences with any category set. Since the inference principles in Section 4 impose connections between predictive inferences for different category sets, we now see that we can interpret these inference principles—or rather, represent them mathematically—as properties of, or restrictions on, coherent inference systems.

7 Representation insensitivity and specificity under exchangeability

Let us now investigate what form the inference principles of representation insensitivity (RI2) and specificity (SP2) take for predictive inference under exchangeability, when such inference can be completely characterised by Bernstein coherent sets of polynomials. This will allow us to reformulate these principles as constraints on—or properties of—inference systems.

Recalling the notations and assumptions in Section 4, we start by considering the surjective (onto) map $C_\rho\colon \mathbb{R}^A \to \mathbb{R}^B$, defined by $C_\rho(\boldsymbol{\alpha})_z := \sum_{x \in A\colon \rho(x)=z} \alpha_x$ for all $\boldsymbol{\alpha} \in \mathbb{R}^A$ and all $z \in B$. It allows us to give the following elegant characterisation of representation insensitivity.

Theorem 2. *An inference system Φ is representation insensitive if and only if for all category sets A and B such that there is an onto map $\rho\colon A \to B$, for all $p \in \mathscr{V}(B)$ and all $\boldsymbol{m} \in \mathscr{N}_A \cup \{\mathbf{0}\}$: $(p \circ C_\rho)\mathrm{B}_{A,\boldsymbol{m}} \in \Phi(A) \Leftrightarrow p\mathrm{B}_{B,C_\rho(\boldsymbol{m})} \in \Phi(B)$.*

Next, we turn to specificity. Let us define the surjective map $r_B\colon \mathbb{R}^A \to \mathbb{R}^B$ by: $r_B(\boldsymbol{\alpha})_z := \alpha_z$ for all $\boldsymbol{\alpha} \in \mathbb{R}^A$ and all $z \in B$. So in particular, $r_B(\boldsymbol{m})$ is the count vector on B obtained by restricting to B the (indices of the) components of the count vector $\boldsymbol{m}$ on A. We also define the one-to-one map $i_A\colon \mathbb{R}^B \to \mathbb{R}^A$ by $i_A(\boldsymbol{\alpha})_x := \alpha_x$ if $x \in B$ and 0 otherwise, for all $\boldsymbol{\alpha} \in \mathbb{R}^B$ and all $x \in A$. This map can be used to define the following one-to-one maps ${}^r\mathrm{I}_{B,A}\colon \mathscr{V}(B) \to \mathscr{V}(A)$, for any $r \in \mathbb{N}_0$, as follows:

$${}^r\mathrm{I}_{B,A}(p) := \sum_{\boldsymbol{n} \in \mathscr{N}_B^{\deg(p)+r}} b_p^{\deg(p)+r}(\boldsymbol{n})\mathrm{B}_{A,i_A(\boldsymbol{n})} \text{ for all polynomials } p \text{ in } \mathscr{V}(B). \quad (13)$$

The maps ${}^r\mathrm{I}_{B,A}$ allow us to give the following elegant characterisation of specificity:

Theorem 3. *An inference system Φ is specific if and only if for all category sets A and B such that $B \subseteq A$, for all $p \in \mathscr{V}(B)$, all $\boldsymbol{m} \in \mathscr{N}_A \cup \{\boldsymbol{0}\}$ and all $r \in \mathbb{N}_0$: ${}^r\mathrm{I}_{B,A}(p)\mathrm{B}_{A,\boldsymbol{m}} \in \Phi(A) \Leftrightarrow p\mathrm{B}_{B,r_B(\boldsymbol{m})} \in \Phi(B)$.*

8 The vacuous inference system

In this and the following sections, we provide explicit and interesting examples of representation insensitive and specific inference systems. We begin with the simplest one: the vacuous inference system Φ_{V}, which is the smallest, or most conservative, coherent inference system. It associates with any category set A the smallest Bernstein coherent set $\Phi_{\mathrm{V}}(A) = \mathscr{H}_{V,A} := \mathscr{V}^+(A)$ containing all the Bernstein positive polynomials—the ones that are guaranteed to be there anyway, by Bernstein coherence alone. Since $\mathscr{V}^+(A)$ consists of all the polynomials that are positive on $\mathrm{int}(\Sigma_A)$ we easily derive that, for any $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\boldsymbol{0}\}$, $\mathscr{H}_{V,A}\rfloor\check{\boldsymbol{m}} = \mathscr{H}_{V,A} = \mathscr{V}^+(A)$. The predictive models for this inference system are now straightforward to find, as they follow directly from Equations (11) and (12). For any $\hat{n} \in \mathbb{N}$ and any $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\boldsymbol{0}\}$, we find that

$$\mathscr{D}^{\hat{n}}_{\mathrm{V},A} = \mathscr{D}^{\hat{n}}_{\mathrm{V},A}\rfloor\check{\boldsymbol{m}} = \left\{f \in \mathscr{G}(A^{\hat{n}})\colon\ \mathrm{Mn}^{\hat{n}}_A(f) \in \mathscr{V}^+(A)\right\} \tag{14}$$

$$\underline{P}^{\hat{n}}_{\mathrm{V},A}(f) = \underline{P}^{\hat{n}}_{\mathrm{V},A}(f|\check{\boldsymbol{m}}) = \min_{\boldsymbol{\theta}\in\Sigma_A} \mathrm{Mn}^{\hat{n}}_A(f|\boldsymbol{\theta}) \text{ for all } f \in \mathscr{G}(A^{\hat{n}}). \tag{15}$$

In particular, $\mathscr{D}^1_{\mathrm{V},A} = \mathscr{D}^1_{\mathrm{V},A}\rfloor\check{\boldsymbol{m}} = \mathscr{G}_{>0}(A)$, and $\underline{P}^1_{\mathrm{V},A}(f) = \underline{P}^1_{\mathrm{V},A}(f|\check{\boldsymbol{m}}) = \min f$ for all $f \in \mathscr{G}(A)$. These are the most conservative exchangeable predictive models there are, and they arise from making no other assessments than exchangeability alone. They are not very interesting, because they involve no non-trivial commitments, and they do not allow learning from observations.

Even though it makes no non-trivial inferences, the vacuous inference system satisfies representation insensitivity and specificity.

Theorem 4. *The vacuous inference system Φ_{V} is coherent, representation insensitive and specific.*

We now show that there is, besides Φ_{V}, an infinity of other, more committal, specific and representation insensitive coherent inference systems.

9 The IDMM inference systems

Imprecise Dirichlet Models (or IDMs, for short) are a family of parametric inference models introduced by Walley [26] as conveniently chosen sets of *Dirichlet densities* $\mathrm{di}_A(\cdot|\boldsymbol{\alpha})$ with constant prior weight s:

$$\{\mathrm{di}_A(\cdot|\boldsymbol{\alpha})\colon\ \boldsymbol{\alpha} \in \mathrm{K}^s_A\}, \text{ with } \mathrm{K}^s_A := \left\{\boldsymbol{\alpha} \in \mathbb{R}^A_{>0}\colon\ \alpha_A = s\right\} = \{s\boldsymbol{t}\colon\ \boldsymbol{t} \in \mathrm{int}(\Sigma_A)\}, \tag{16}$$

for any value of the (so-called) hyperparameter $s \in \mathbb{R}_{>0}$ and any category set A. The Dirichlet densities $\mathrm{di}_A(\cdot|\boldsymbol{\alpha})$ are defined on $\mathrm{int}(\Sigma_A)$.

These IDMs generalise the Imprecise Beta models introduced earlier by Walley [25]. In a later paper [29], Walley and Bernard introduced a closely related family of predictive inference models, called the Imprecise Dirichlet Multinomial Models (or IDMMs, for short). We use the ideas behind Walley's IDM(M)s to construct an interesting family of coherent inference systems. Interestingly, we shall need a slightly modified version of Walley's IDMs to make things work. The reason for this is that Walley's original version, as described by Equation (16), has a number of less desirable properties, that were either unknown to, or ignored by, Walley and Bernard. For our present purposes, it suffices to mention that, contrary to what is often claimed, and in contradistinction with our new version, inferences using the original version of the IDM(M) do not always become more conservative (or less committal) as the hyperparameter s increases.

In our version, rather than using the hyperparameter sets K_A^s, we consider the sets

$$\Delta_A^s := \left\{\boldsymbol{\alpha} \in \mathbb{R}_{>0}^A \colon \alpha_A < s\right\} \text{ for any } s \in \mathbb{R}_{>0}.$$

Observe that $\Delta_A^s = \{s'\boldsymbol{t} \colon s' \in \mathbb{R}_{>0}, s' < s \text{ and } \boldsymbol{t} \in \mathrm{int}(\Sigma_A)\} = \bigcup_{0<s'<s} \mathrm{K}_A^{s'}$. For any $s \in \mathbb{R}_{>0}$, and any category set A, we now consider the following set of desirable polynomials p, with positive Dirichlet expectation $\mathrm{Di}_A(p|\boldsymbol{\alpha})$ for all hyperparameters $\boldsymbol{\alpha} \in \Delta_A^s$:

$$\mathscr{H}_{\mathrm{IDM},A}^s := \{p \in \mathscr{V}(A) \colon (\forall \boldsymbol{\alpha} \in \Delta_A^s)\, \mathrm{Di}_A(p|\boldsymbol{\alpha}) > 0\}.$$

We shall see further on in Theorem 5 that this set is Bernstein coherent. We call the inference system Φ_{IDM}^s, defined by $\Phi_{\mathrm{IDM}}^s(A) := \mathscr{H}_{\mathrm{IDM},A}^s$ for all category sets A, the *IDMM inference system* with hyperparameter $s > 0$. The corresponding updated models are, for any $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\mathbf{0}\}$, given by:

$$\mathscr{H}_{\mathrm{IDM},A}^s \rfloor \check{\boldsymbol{m}} = \{p \in \mathscr{V}(A) \colon (\forall \boldsymbol{\alpha} \in \Delta_A^s)\, \mathrm{Di}_A(p|\check{\boldsymbol{m}} + \boldsymbol{\alpha}) > 0\} \tag{17}$$

Using these expressions, the predictive models for the IDMM inference system are straightforward to find; it suffices to apply Equations (11) and (12). For any $\hat{n} \in \mathbb{N}$ and any $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\mathbf{0}\}$:

$$\mathscr{D}_{\mathrm{IDM},A}^{s,\hat{n}} \rfloor \check{\boldsymbol{m}} = \left\{f \in \mathscr{G}(A^{\hat{n}}) \colon (\forall \boldsymbol{\alpha} \in \Delta_A^s)\, \mathrm{Di}_A(\mathrm{Mn}_A^{\hat{n}}(f)|\check{\boldsymbol{m}} + \boldsymbol{\alpha}) > 0\right\}, \tag{18}$$

$$\underline{P}_{\mathrm{IDM},A}^{s,\hat{n}}(f|\check{\boldsymbol{m}}) = \inf_{\boldsymbol{\alpha} \in \Delta_A^s} \mathrm{Di}_A(\mathrm{Mn}_A^{\hat{n}}(f)|\check{\boldsymbol{m}} + \boldsymbol{\alpha}) \text{ for all } f \in \mathscr{G}(A^{\hat{n}}), \tag{19}$$

where:

$$\mathrm{Di}_A(\mathrm{Mn}_A^{\hat{n}}(f)|\check{\boldsymbol{m}} + \boldsymbol{\alpha}) = \sum_{\hat{\boldsymbol{m}} \in \mathscr{N}_A^{\hat{n}}} \mathrm{Hy}_A^{\hat{n}}(f|\hat{\boldsymbol{m}}) \frac{1}{(\check{m}_A + \alpha_A)^{(\hat{n})}} \binom{\hat{n}}{\hat{\boldsymbol{m}}} \prod_{x \in A} (\check{m}_x + \alpha_x)^{(\hat{m}_x)}.$$

In general, these expressions seem forbidding, but for $\hat{n} = 1$, the so-called *immediate prediction models* are manageable enough: for any $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\mathbf{0}\}$

$$\mathscr{D}^{s,1}_{\mathrm{IDM},A}\rfloor\check{\boldsymbol{m}} = \left\{ f \in \mathscr{G}(A) \colon f > -\frac{1}{s}\sum_{x\in A} f(x)\check{m}_x \right\}, \tag{20}$$

$$\underline{P}^{s,1}_{\mathrm{IDM},A}(f|\check{\boldsymbol{m}}) = \frac{1}{\check{m}_A + s}\sum_{x\in A} f(x)\check{m}_x + \frac{s}{\check{m}_A + s}\min f \text{ for all } f \in \mathscr{G}(A), \tag{21}$$

Interestingly, the immediate prediction models of our version of the IDMM inference system coincide with those of Walley's original version.

The IDMM inference systems constitute an uncountably infinite family of coherent inference systems, each of which satisfies the representation insensitivity and specificity requirements.

Theorem 5. *For any $s \in \mathbb{R}_{>0}$, the IDMM inference system Φ^s_{IDM} is coherent, representation insensitive and specific.*

10 The Haldane inference system

We can ask ourselves whether there are representation insensitive (and specific) inference systems whose *posterior* predictive lower previsions become precise (linear) previsions. In the present section, we show that this is indeed the case. We use the family of IDMM inference systems Φ^s_{IDM}, $s \in \mathbb{R}_{>0}$, to define an inference system Φ_{H} that is more committal than each of them:

$$\Phi_{\mathrm{H}}(A) = \mathscr{H}_{\mathrm{H},A} := \bigcup_{s\in\mathbb{R}_{>0}} \mathscr{H}^s_{\mathrm{IDM},A} = \bigcup_{s\in\mathbb{R}_{>0}} \Phi^s_{\mathrm{IDM}}(A) \text{ for all category sets } A.$$

We call this Φ_{H} the *Haldane inference system*, for reasons that will become clear further on in this section.

Theorem 6. *The Haldane inference system Φ_{H} is coherent, representation insensitive and specific.*

It can be shown that, due to its representation insensitivity, the Haldane system satisfies prior near-ignorance: this means that before making any observation, its immediate prediction model is vacuous, and as far away from a precise probability model as possible. But after making even a single observation, its inferences become precise-probabilistic: they coincide with the inferences generated by the Haldane (improper) prior. To get there, we first take a look at the models involving sets of desirable gambles. For any $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\boldsymbol{0}\}$:

$$\mathscr{H}_{\mathrm{H},A}\rfloor\check{\boldsymbol{m}} = \{p \in \mathscr{V}(A) \colon (\exists s \in \mathbb{R}_{>0})(\forall \boldsymbol{\alpha} \in \Delta^s_A)\,\mathrm{Di}_A(p|\check{\boldsymbol{m}} + \boldsymbol{\alpha}) > 0\}. \tag{22}$$

The corresponding predictive models are easily derived by applying Equation (11). For any $\hat{n} \in \mathbb{N}$ and any $\check{\boldsymbol{m}} \in \mathscr{N}_A \cup \{\boldsymbol{0}\}$:

$$\mathscr{D}^{\hat{n}}_{\mathrm{H},A}\rfloor\check{\boldsymbol{m}} = \left\{ f \in \mathscr{G}(A^{\hat{n}}) \colon (\exists s \in \mathbb{R}_{>0})(\forall \boldsymbol{\alpha} \in \Delta^s_A)\,\mathrm{Di}_A(\mathrm{Mn}^{\hat{n}}_A(f)|\check{\boldsymbol{m}} + \boldsymbol{\alpha}) > 0 \right\}. \tag{23}$$

The immediate prediction models are obtained by combining Equations (23), (18) and (20). For any $\check{\boldsymbol{m}} \in \mathscr{N}_A$:

$$\mathscr{D}^1_{\mathrm{H},A} = \mathscr{G}_{>0}(A) \text{ and } \mathscr{D}^1_{\mathrm{H},A}\rfloor\check{\boldsymbol{m}} = \left\{ f \in \mathscr{G}(A) \colon \sum_{x\in A} f(x)\check{m}_x > 0 \right\} \cup \mathscr{G}_{>0}(A). \quad (24)$$

It turns out that the expressions for the corresponding lower previsions are much more manageable. In particular, for $\check{\boldsymbol{m}} = \mathbf{0}$:

$$\underline{P}^{\hat{n}}_{\mathrm{H},A}(f) = \min_{x\in A} f(x,x,\dots,x) \text{ for all } f \in \mathscr{G}(A^{\hat{n}}), \quad (25)$$

and for any $\check{\boldsymbol{m}} \in \mathscr{N}_A$:

$$\underline{P}^{\hat{n}}_{\mathrm{H},A}(f|\check{\boldsymbol{m}}) = \overline{P}^{\hat{n}}_{\mathrm{H},A}(f|\check{\boldsymbol{m}}) = P^{\hat{n}}_{\mathrm{H},A}(f|\check{\boldsymbol{m}}) = \sum_{\boldsymbol{n}\in\mathscr{N}^{\hat{n}}_A} \mathrm{Hy}^{\hat{n}}_A(f|\boldsymbol{n}) \binom{\hat{n}}{\boldsymbol{n}} \frac{\prod_{x\in A} \check{m}_x^{(n_x)}}{\check{m}_A^{(\hat{n})}}. \quad (26)$$

For the immediate prediction models, we find that for any $\check{\boldsymbol{m}} \in \mathscr{N}_A$:

$$\underline{P}^1_{\mathrm{H},A}(f) = \min f \text{ and } P^1_{\mathrm{H},A}(f|\check{\boldsymbol{m}}) = \sum_{x\in A} f(x)\frac{\check{m}_x}{\check{m}_A} \text{ for all } f \in \mathscr{G}(A), \quad (27)$$

The precise posterior predictive previsions in Equation (26) are exactly the ones that would be found were we to formally apply Bayes's rule with a multinomial likelihood and *Haldane's improper prior* [14, 16, 15], whose 'density' is a function on $\mathrm{int}(\Sigma_A)$ proportional to $\prod_{x\in A}\theta_x^{-1}$. This, of course, is why we use Haldane's name for the inference system that produces them. Our argumentation shows that there is nothing wrong with these posterior predictive previsions, as they are based on coherent inferences. In fact, our analysis shows that there is an infinity of *precise and proper* priors on the simplex Σ_A that, together with the multinomial likelihood, are coherent with these posterior predictive previsions: every linear prevision on $\mathscr{V}(A)$ that dominates the coherent lower prevision $\underline{H}_{\mathrm{H},A}$ on $\mathscr{V}(A)$,[5,6] as defined by $\underline{H}_{\mathrm{H},A}(p) := \sup\{\mu \in \mathbb{R} \colon p - \mu \in \mathscr{H}_{\mathrm{H},A}\}$ for all polynomials p on Σ_A.

11 Conclusion

We believe this is the first paper that tries to deal in a systematic fashion with predictive inference under exchangeability using imprecise probability models. A salient feature of our approach is that we consistently use coherent sets of desirable gambles

[5] Actually, a suitably adapted version of coherence, where the gambles are restricted to the polynomials on Σ_A.

[6] It is an immediate consequence of the F. Riesz Extension Theorem that each such linear prevision is the restriction to polynomials of the expectation operator of some unique σ-additive probability measure on the Borel sets of Σ_A; see for instance Ref. [5].

as our uncertainty models of choice. This allows us, in contradistinction with most other approaches in probability theory, to avoid problems with determining unique conditional models from unconditional ones when conditioning on events with (lower) probability zero. A set of polynomials $\mathscr{H}_A$ completely determines all prior and posterior predictive models $\mathscr{D}^{\hat{n}}_A\rfloor\check{\boldsymbol{m}}$ and $\underline{P}^{\hat{n}}_A(\cdot|\check{\boldsymbol{m}})$, even when the (lower) prior probability $\underline{P}^{\check{n}}_A([\check{\boldsymbol{m}}])=\underline{H}_A(\mathrm{B}_{A,\check{\boldsymbol{m}}})$ of observing the count vector $\check{\boldsymbol{m}}$ is zero. An approach using only lower previsions and probabilities would make this much more complicated and involved, if not impossible. Indeed, it can be proved that any inference system that satisfies representation insensitivity has near-vacuous prior predictive models, and that therefore its prior predictive lower previsions must satisfy $\underline{P}^{\check{n}}_A([\check{\boldsymbol{m}}])=0$. This simply means that it is *impossible* in a representation insensitive inference system for the prior lower previsions to uniquely determine posteriors. And therefore any systematic way of dealing with such inference systems must be able to resolve—or deal with—this non-unicity in some way. We believe our approach involving coherent sets of desirable gambles is one of the mathematically most elegant ways of doing this.

We might also wonder whether there are other representation insensitive and specific inference systems. We suggest, as candidates for further consideration, the inference systems that can be derived using Walley's bounded derivative model [27], and inference systems that can be constructed using sets of infinitely divisible distributions, as recently proposed by Mangili and Benavoli [20].

Acknowledgements

Gert de Cooman's research was partially funded through project number 3G012512 of the Research Foundation Flanders (FWO). Jasper De Bock is a PhD Fellow of the Research Foundation Flanders and wishes to acknowledge its financial support. Marcio Diniz was supported by FAPESP (São Paulo Research Foundation), under the project 2012/14764-0 and wishes to thank the SYSTeMS Research Group at Ghent University for its hospitality and support during his sabbatical visit there.

References

1. Augustin, T., Coolen, F.P.A., de Cooman, G., Troffaes, M.C.M. (eds.): Introduction to Imprecise Probabilities. John Wiley & Sons (2014)
2. Bernard, J.M.: Bayesian analysis of tree-structured categorized data. Revue Internationale de Systémique **11**, 11–29 (1997)
3. Bernard, J.M.: An introduction to the imprecise Dirichlet model for multinomial data. International Journal of Approximate Reasoning **39**, 123–150 (2005)
4. Cifarelli, D.M., Regazzini, E.: De Finetti's contributions to probability and statistics. Statistical Science **11**, 253–282 (1996)
5. de Cooman, G., Miranda, E.: The F. Riesz Representation Theorem and finite additivity. In: D. Dubois, M.A. Lubiano, H. Prade, M.A. Gil, P. Grzegorzewski, O. Hryniewicz (eds.) Soft

Methods for Handling Variability and Imprecision (Proceedings of SMPS 2008), pp. 243–252. Springer (2008)

6. de Cooman, G., Miranda, E.: Irrelevant and independent natural extension for sets of desirable gambles. Journal of Artificial Intelligence Research **45**, 601–640 (2012). URL http://www.jair.org/vol/vol45.html
7. de Cooman, G., Miranda, E., Quaeghebeur, E.: Representation insensitivity in immediate prediction under exchangeability. International Journal of Approximate Reasoning **50**(2), 204–216 (2009). DOI 10.1016/j.ijar.2008.03.010
8. de Cooman, G., Quaeghebeur, E.: Exchangeability and sets of desirable gambles. International Journal of Approximate Reasoning **53**(3), 363–395 (2012). Special issue in honour of Henry E. Kyburg, Jr.
9. de Cooman, G., Quaeghebeur, E., Miranda, E.: Exchangeable lower previsions. Bernoulli **15**(3), 721–735 (2009). DOI 10.3150/09-BEJ182. URL http://hdl.handle.net/1854/LU-498518
10. Couso, I., Moral, S.: Sets of desirable gambles: conditioning, representation, and precise probabilities. International Journal of Approximate Reasoning **52**(7), 1034–1055 (2011)
11. de Finetti, B.: La prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré **7**, 1–68 (1937). English translation in [18]
12. de Finetti, B.: Teoria delle Probabilità. Einaudi, Turin (1970)
13. de Finetti, B.: Theory of Probability: A Critical Introductory Treatment. John Wiley & Sons, Chichester (1974–1975). English translation of [12], two volumes
14. Haldane, J.B.S.: On a method of estimating frequencies. Biometrika **33**, 222–225 (1945)
15. Jaynes, E.T.: Probability Theory: The Logic of Science. Cambridge University Press (2003)
16. Jeffreys, H.: Theory of Probability. Oxford Classics series. Oxford University Press (1998). Reprint of the third edition (1961), with corrections
17. Johnson, N.L., Kotz, S., Balakrishnan, N.: Discrete Multivariate Distributions. Wiley Series in Probability and Statistics. John Wiley and Sons, New York (1997)
18. Kyburg Jr., H.E., Smokler, H.E. (eds.): Studies in Subjective Probability. Wiley, New York (1964). Second edition (with new material) 1980
19. Lad, F.: Operational Subjective Statistical Methods: A Mathematical, Philosophical and Historical Introduction. John Wiley & Sons (1996)
20. Mangili, F., Benavoli, A.: New prior near-ignorance models on the simplex. In: F. Cozman, T. Denœux, S. Destercke, T. Seidenfeld (eds.) ISIPTA '13 – Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications, pp. 213–222. SIPTA (2013)
21. Moral, S.: Epistemic irrelevance on sets of desirable gambles. Annals of Mathematics and Artificial Intelligence **45**, 197–214 (2005). DOI 10.1007/s10472-005-9011-0
22. Quaeghebeur, E.: Introduction to Imprecise Probabilities, chap. Desirability. John Wiley & Sons (2014)
23. Quaeghebeur, E., de Cooman, G., Hermans, F.: Accept & reject statement-based uncertainty models. International Journal of Approximate Reasoning (2013). Submitted for publication
24. Rouanet, H., Lecoutre, B.: Specific inference in ANOVA: From significance tests to Bayesian procedures. British Journal of Mathematical and Statistical Psychology **36**(2), 252–268 (1983). DOI 10.1111/j.2044-8317.1983.tb01131.x
25. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
26. Walley, P.: Inferences from multinomial data: learning about a bag of marbles. Journal of the Royal Statistical Society, Series B **58**, 3–57 (1996). With discussion
27. Walley, P.: A bounded derivative model for prior ignorance about a real-valued parameter. Scandinavian Journal of Statistics **24**(4), 463–483 (1997). DOI 10.1111/1467-9469.00075
28. Walley, P.: Towards a unified theory of imprecise probability. International Journal of Approximate Reasoning **24**, 125–148 (2000)
29. Walley, P., Bernard, J.M.: Imprecise probabilistic prediction for categorical data. Tech. Rep. CAF-9901, Laboratoire Cognition et Activitées Finalisées, Université de Paris 8 (1999)
30. Williams, P.M.: Indeterminate probabilities. In: M. Przelecki, K. Szaniawski, R. Wojcicki (eds.) Formal Methods in the Methodology of Empirical Sciences, pp. 229–246. Reidel, Dordrecht (1976). Proceedings of a 1974 conference held in Warsaw