Coherent Predictive Inference under Exchangeability with Imprecise Probabilities (Extended Abstract)[*]

Gert de Cooman[*], **Jasper De Bock**[*] and **Márcio Alves Diniz**[†]

[*]IDLab, Ghent University, Belgium

[†]Department of Statistics, Federal University of São Carlos, Brazil

{gert.decooman,jasper.debock}@ugent.be and marcio.alves.diniz@gmail.com

Abstract

Coherent reasoning under uncertainty can be represented in a very general manner by coherent sets of desirable gambles. This leads to a more general foundation for coherent (imprecise-)probabilistic inference that allows for indecision. 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. We define an inference system as a map that associates a Bernstein coherent cone of polynomials with every finite category set. Inference principles can then be represented mathematically as restrictions on such maps, which allows us to develop a notion of conservative inference under such inference principles. We discuss, as particular examples, representation insensitivity and specificity, and show that there is an infinity of inference systems that satisfy these two principles.

1 The Setting: Predictive Inference

We deal with predictive inference for categorical variables, and 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 $X_1 = x_1, X_2 = x_2, \ldots, X_{\check{n}} = x_{\check{n}}$, we consider some subject's belief model for the next $\hat{n} \geq 1$ variables $X_{\check{n}+1}, \ldots X_{\check{n}+\hat{n}}$. In the probabilistic tradition that we build on in the paper, this belief is modelled by a conditional predictive probability mass function $p^{\hat{n}}(\cdot|x_1, \ldots, x_{\check{n}})$ on the set $A^{\hat{n}}$ of their possible values. These probability mass functions can be used for prediction or estimation, for statistical inferences, and in decision making. In this sense, predictive inference lies at the heart of statistics, and more generally, of learning under uncertainty. For this reason, it is also of crucial importance for dealing with uncertainty in Artificial Intelligence. We refer to the synthesis by Geisser [1993] and the collection of essays by Zabell [2005] for introductions to predictive inference and the underlying issues that the paper is also concerned with.

---

[*]This paper is an extended abstract of an article in the Journal of Artificial Intelligence Research [De Cooman *et al.*, 2015].

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

A common assumption about the variables X_n is that they are *exchangeable*, meaning roughly that the subject believes that the order in which they are observed, or present themselves, has no influence on the decisions and inferences she will make regarding these variables. This assumption, and the analysis of its consequences, goes back to de Finetti [1937]. His famous Representation Theorem states, in essence, that the time-consistent and coherent conditional and unconditional predictive probability mass functions associated with a countably infinite exchangeable sequence of variables in A are characterised by, and characterise, a unique probability measure on the Borel sets of the simplex of all probability mass functions on A, called their *representation*.[1]

2 The Central Problem of Predictive Inference

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 might answer that this question needs no answer: a subject's personal predictive probabilities are entirely hers, and time consistency and coherence are the only requirements she should heed. Earlier scholars, like Laplace and Bayes, whom we would now also call subjectivists, invoked the Principle of Indifference

---

[1]To clarify the connection with our argumentation in the paper, the essence of de Finetti's argument is that the representation is a coherent prevision on the set of all multinomial polynomials on this simplex [De Cooman *et al.*, 2009b]. As a (*finitely* additive) coherent prevision, it can be extended uniquely only so far as to the set of all lower semicontinuous functions, but it does determine a unique (*countably* additive) probability measure on the Borel sets of that simplex, through the F. Riesz Representation Theorem [De Cooman and Miranda, 2008; Troffaes and De Cooman, 2014].

to justify using a specific class of predictive mass functions. 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. The logicians W. E. Johnson [1924] and, in a much more systematic fashion, Rudolf Carnap [1952] tried to develop an axiom system for predictive inference based on such reasonable inference principles. Carnap's first group of axioms is related to what we have called coherence, but as we suggested, these by themselves are too weak to single out a particular predictive model. His second group consists of invariance axioms, including exchangeability. He also included an axiom of instantial relevance, translating the intuitive principle that predictive inferences should actually learn from experience. His last axiom, predictive irrelevance, was also proposed earlier by Johnson and called the *sufficientness postulate* by Good [1965]. Armed with these axioms, Carnap was able to derive a continuum of probabilistic inference rules, closely related to the Dirichlet multinomial model and to the Imprecise Dirichlet Multinomial Model (IDMM) proposed by Walley [1996] and Walley and Bernard [1999].

Our point of view holds the middle ground between the subjectivist and logicist positions: it ought to be possible for a subject to make assessments for certain predictive probabilities, and to combine these with certain inference principles she finds reasonable, or which suit her purpose for the problem at hand. Indeed, the inference systems we introduce and discuss in the paper, and the notion of conservative coherent inference we associate with them, provide an elegant framework and tools for making conservative coherent predictive inferences that combine (local) subjective probability assessments with (general) inference principles. And our discussion later in the paper on characterising the immediate predictions for the IDMM constitutes an exercise in precisely that.

3 Conservative Probabilistic Inference

This idea of *conservative probabilistic inference* brings us to what we believe is our main contribution. It is a central idea in de Finetti's [1974] approach to probability that when a subject makes probability assessments, we can consider them as bounds on so-called precise probability models. Calculating such most conservative but tightest bounds is indeed what de Finetti's [1974] Fundamental Theorem of Prevision (see also [Lad, 1996]) is about. The theory of imprecise probabilities, brought to a synthesis by Williams [1976] and Walley [1991; 2000], looks at conservative probabilistic inference in precisely this way: how can we calculate efficiently the consequences—in the sense of most conservative tightest bounds—of making certain probability assessments. These may be local assessments, such as inequalities imposed on the probabilities or previsions of certain events or variables, or structural assessments, such as independence, or exchangeability.

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. Another, related, advantage is that they allow for *indecision* to be modelled explicitly: loosely stated, if the imposed bounds on probabilities allow for more than one probability model as a solution, it may very well be that of two actions, the first has the higher expected utility for one compatible probability model, and the smaller for another compatible probability model, meaning that neither action is robustly preferred over the other. So with this current stated model for her beliefs, a subject is then undecided between these actions. In the paper, we give a concise overview of the relevant ideas, models and techniques in the field of imprecise probabilities. A much more extensive and detailed recent overview of this area of research was published by Augustin *et al.* [2014].

4 So What Do We Aim at, Then?

Our 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—conservative coherent predictive inference using imprecise probability models. Using such models also allows 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 that theories of learning under uncertainty in general, and predictive inference in particular, at least allow us (i) to start out with conservative, very imprecise and indecisive models when little has been learned, and (ii) to become more precise and decisive as more observations come in. Our abstract notion of an inference system allows for—but does not force—such behaviour, and we provide examples of concrete inference systems that display it.

Our work builds on, but manages to reach much further than, an earlier paper by one of the authors [De Cooman *et al.*, 2009a]. One reason why it does, is that this earlier work deals only with immediate prediction models—which focus on the case $\hat{n} = 1$—and as we argue extensively in the paper, *predictive inference using imprecise probabilities is not completely determined by immediate prediction*, contrary to what we can expect when using precise probabilities. But the main reason is that we are now in a position to use a very powerful mathematical language to represent imprecise-probabilistic inferences: Walley's [2000] coherent sets of desirable gambles. Earlier imprecise probability models [Boole, 1952; 1847; Koopman, 1940] centred on lower and upper probability bounds for events—or propositions. Later on [Walley, 1991, Section 2.7], it became apparent that this language of events and *lower* and *upper probabilities* is lacking in power of expression: a much more expressive theory uses random variables and their *lower previsions* or *expectations*. This successful theory of coherent lower previsions is by now quite well developed [Walley, 1991; Augustin *et al.*, 2014; Troffaes and De Cooman, 2014]. But it faces a number of problems, such as its mathematical as well as conceptual complexity, especially when dealing with conditioning and independence, and the fact that, as is the case with many other approaches to probability, and as we show in the paper, it has issues with conditioning on sets of (lower) probability zero.

A very attractive solution to these problems was offered by Walley [2000], in the form of coherent sets of desirable gambles, inspired by earlier ideas [Smith, 1961; Williams, 1975; Seidenfeld *et al.*, 1995]. Here, the primitive notions are not

probabilities of events, nor expectations of random variables. The focus is rather on 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*. Of course, stating that a gamble is desirable also leads to a particular lower prevision assessment: it provides a lower bound of zero on the prevision of the gamble. We explain in great detail in the paper why we prefer to use sets of desirable gambles as our basic uncertainty models.

5 A Walk through the Paper

In summary, then, our aim is to use sets of desirable gambles to extend the existing probabilistic theory of predictive inference. How do we go about doing this?

5.1 Setting up the Machinery ...

We first introduce basic building blocks, and give an overview of relevant notions and results concerning our imprecise probability model of choice—coherent sets of desirable gambles. 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.

Next, 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 $p^{\hat{n}}(\cdot|x_1, \ldots, x_{\tilde{n}})$, and they constitute the basic tools we work with. We also explain what are the proper formulations for the above-mentioned time consistency and coherence requirements in this more general context.

We then take a first important step: 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. Pooling invariance—or what Walley [1996] has called the Representation Invariance Principle (RIP)—and renaming invariance seem reasonable requirements for any type of predictive inference, and category permutation invariance seems a natural thing to require when starting from a state of complete ignorance. Taken together, they constitute what we call *representation insensitivity*. 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. Another inference principle we look at imposes the so-called *specificity* property: when predictive inference is specific, then for a certain type of 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 [Bernard, 1997].

As a next important step, we recall from the literature [De Cooman *et al.*, 2009b; De Cooman and Quaeghebeur, 2012] how to deal with exchangeability when our predictive inference models are imprecise, and show that de Finetti's Repre-

sentation Theorem can be significantly generalised. In particular, the time-consistent and coherent predictive sets of desirable gambles can be characterised completely by a set of (multivariate) polynomials on the simplex of all probability mass functions on the category set.[2] This set of polynomials must satisfy a number of properties, which taken together define the notion of *Bernstein coherence*. Without becoming too technical at this point, our conclusion is that, in our more general context, the representing probability measure on the simplex of all probability mass functions is replaced by a Bernstein coherent set of polynomials on this simplex. This set of polynomials serves completely the same purpose as the representing probability measure: it completely determines, and conveniently and densely summarises, all predictive inferences. This is why all further developments in the paper are expressed in terms of such Bernstein coherent sets of polynomials.

We next introduce coherent inference systems 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. So a coherent inference system is a way of fixing completely all coherent predictive inferences for all possible category sets. Our reasons for introducing such coherent inference systems are twofold. First of all, the above-mentioned inference principles impose connections between predictive inferences for different category sets, so we can represent such inference principles mathematically as restrictions on coherent inference systems. Secondly, it allows us to extend our method of conservative inference to also take into account principles for predictive inference, or more generally, predictive inference for multiple category sets at once. This leads to a method of combining (local) predictive probability assessments with (global) inference principles to produce the most conservative predictive inferences compatible with them.

5.2 ... And Getting It to Work

As a first illustration of the power of our methodology, we then look at immediate prediction: what implications do representation insensitivity and specificity have for predictive inference about the single next observation? Our approach allows us to streamline, simplify and significantly extend previous attempts in this direction by De Cooman *et al.* [2009a].

Then follow several explicit examples, worked out in great detail, to show that there are quite a few different types—even uncountable infinities—of coherent inference systems that are representation insensitive and/or specific. We discuss the vacuous and nearly vacuous inference systems, the skeptically cautious inference system, the family of IDMM inference systems, the family of skeptical IDMM inference systems, and the Haldane inference system. Most of these inference systems, apart from the IDMM, appear in the paper for the first time. Also, we believe that we are the first to publish a detailed and explicit—as well as still elegant—proof that the IDMM inference systems are indeed representation insensitive and specific. It should already be mentioned here, however, that our IDMM inference systems are based on a modified, and arguably better behaved, version of the models originally introduced by

---

[2]In contrast with de Finetti's version, our version has no problems with conditioning on observed sequences of (lower) probability zero.

Walley and Bernard [Walley, 1996; Walley and Bernard, 1999; Bernard, 2005]; we show that the original IDMM is not specific and that, contrary to what is often claimed, it does not satisfy the so-called nestedness property.

Our results also disprove the conjecture [Bernard, 2007; De Cooman *et al.*, 2009a] that the IDMM inference systems—our version or the original one—are the only ones, or even the most conservative ones, that satisfy both representation insensitivity and specificity. But we do show that the IDMM family of immediate predictions—which are the same for our version and for the original one—are in a definite sense the most conservative ones that are representation insensitive and specific, and satisfy another requirement, which we have called 'having concave surprise'.

In the conclusions of the paper, we point to a number of surprising consequences of our results, and discuss avenues for further research.

Acknowledgments

References

[Augustin *et al.*, 2014] Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[Bernard, 1997] Jean-Marc Bernard. Bayesian analysis of tree-structured categorized data. *Revue Internationale de Systémique*, 11:11–29, 1997.

[Bernard, 2005] Jean-Marc Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39:123–150, 2005.

[Bernard, 2007] Jean-Marc Bernard. In personal conversation. 2007.

[Boole, 1847] George Boole. *The Laws of Thought*. Dover Publications, New York, 1847, reprinted in 1961.

[Boole, 1952] George Boole. *Studies in Logic and Probability*. Dover Publications, Mineola, NY, 2004, reprint of the work originally published by Watts & Co., London, in 1952.

[Carnap, 1952] Rudolf Carnap. *The continuum of inductive methods*. The University of Chicago Press, 1952.

[De Cooman and Miranda, 2008] Gert de Cooman and Enrique Miranda. The F. Riesz Representation Theorem and finite additivity. In *Soft Methods for Handling Variability and Imprecision (Proceedings of SMPS 2008)*, pages 243–252. Springer, 2008.

[De Cooman and Quaeghebeur, 2012] Gert de Cooman and Erik Quaeghebeur. Exchangeability and sets of desirable gambles. *International Journal of Approximate Reasoning*, 53:363–395, 2012. Special issue in honour of Henry E. Kyburg, Jr.

[De Cooman *et al.*, 2009a] Gert de Cooman, Enrique Miranda, and Erik Quaeghebeur. Representation insensitivity in immediate prediction under exchangeability. *International Journal of Approximate Reasoning*, 50:204–216, 2009.

[De Cooman *et al.*, 2009b] Gert de Cooman, Erik Quaeghebeur, and Enrique Miranda. Exchangeable lower previsions. *Bernoulli*, 15:721–735, 2009.

[De Cooman *et al.*, 2015] Gert de Cooman, Jasper De Bock, and Márcio Diniz. Coherent predictive inference under exchangeability with imprecise probabilities. *Journal of Artificial Intelligence Research*, 52:1–95, 2015.

[de Finetti, 1937] Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937.

[de Finetti, 1974] Bruno de Finetti. *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons, Chichester, 1974–1975.

[Geisser, 1993] Seymour Geisser. *Predictive Inference: An Introduction*. Chapman & Hall, 1993.

[Good, 1965] Irving John Good. *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. The MIT Press, 1965.

[Johnson, 1924] William Ernest Johnson. *Logic, Part III. The Logical Foundations of Science*. Cambridge University Press, 1924. Reprinted by Dover Publications in 1964.

[Koopman, 1940] Bernard O. Koopman. The Axioms and Algebra of Intuitive Probability. *The Annals of Mathematics, Second Series*, 41:269–292, 1940.

[Lad, 1996] Frank Lad. *Operational Subjective Statistical Methods: A Mathematical, Philosophical and Historical Introduction*. John Wiley & Sons, 1996.

[Seidenfeld *et al.*, 1995] Teddy Seidenfeld, Mark J. Schervish, and Jay B. Kadane. A representation of partially ordered preferences. *The Annals of Statistics*, 23:2168–2217, 1995.

[Smith, 1961] Cedric A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series A*, 23:1–37, 1961.

[Troffaes and De Cooman, 2014] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, 2014.

[Walley and Bernard, 1999] Peter Walley and Jean-Marc Bernard. Imprecise probabilistic prediction for categorical data. Technical Report CAF-9901, Laboratoire Cognition et Activitées Finalisées, Université de Paris 8, January 1999.

[Walley, 1991] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[Walley, 1996] Peter Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996. With discussion.

[Walley, 2000] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[Williams, 1975] Peter M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975. Revised journal version: [Williams, 2007].

[Williams, 1976] Peter M. Williams. Indeterminate probabilities. In *Formal Methods in the Methodology of Empirical Sciences*, pages 229–246. Reidel, Dordrecht, 1976. Proceedings of a 1974 conference held in Warsaw.

[Williams, 2007] Peter M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383, 2007. Revised journal version of [Williams, 1975].

[Zabell, 2005] Sandy L. Zabell. *Symmetry and Its Discontents: Essays on the History of Inductive Probability*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, Cambridge, UK, 2005.