SPRINGER BRIEFS IN COMPUTER SCIENCE

Dirk Draheim

Generalized Jeffrey Conditionalization

A Frequentist Semantics of Partial Conditionalization

Springer

SpringerBriefs in Computer Science

More information about this series at http://www.springer.com/series/10028

Dirk Draheim

Generalized Jeffrey Conditionalization

A Frequentist Semantics of Partial Conditionalization

Springer

Dirk Draheim
Large-Scale Systems Group
Tallinn University of Technology
Tallinn
Estonia

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-319-69867-0 ISBN 978-3-319-69868-7 (eBook)
https://doi.org/10.1007/978-3-319-69868-7

Library of Congress Control Number: 2017956747

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

In my view, mathematics of the 21st century can be characterized by the attempt at *automating* mathematical reasoning. By Gödel, we know that this will never be completely possible: The higher we go in the sophistication of automation the more remote is the horizon of where we would like to go next. However, it is possible and exciting to conquer higher and higher levels of automation.

The 20th century was the century of *formalizing* mathematical reasoning, which is the first step towards *automating* mathematical reasoning. As a side-product of mathematical formalization, the notion of "universal computer" – which in essence is a mathematical and not an engineering concept – was invented. The enormous impact of this notion on all aspects of science, technology, economy, and society as a whole, by now, is understood by everybody. The impact of *automating* mathematical reasoning (mathematical invention and mathematical verification) will generate bigger and bigger waves of understanding the world and of societal transformation. The waves will include such theoretical areas like, for example, the build-up of web-accessible global and comprehensive mathematical knowledge bases and such practical effects like, for example, deriving hidden knowledge from social media messages.

The level of formalization is not equally high in all areas of mathematics. In this book, Dirk Draheim lays the ground for the formalization of an important part of mathematics that also has high relevance to modern data science: probabilistic reasoning. He clarifies the frequentist semantics of the fundamental notion of partial conditional probability and reveals the subtle differences and the relation between this frequentist and the established Bayesian view. This is the first time that the many results that are due to earlier publications in this area are brought into a coherent form. The concepts can be made operational in today's standard programming paradigms. Thus, the foundational results are immediately available also for practical probabilistic modeling, which is of course of high relevance in current data science and artificial intelligence.

I wish this book wide distribution both in the research community and in the business world.

Research Institute for Symbolic Computation
Johannes Kepler University
Linz/Hagenberg, October 2017 *Bruno Buchberger*

Preface

Statistics is the language of science; however, the semantics of probabilistic reasoning is still a matter of discourse. In this book, I provide a frequentist semantics for conditionalization on partially known events. The resulting frequentist partial (F.P.) conditionalization generalizes Jeffrey conditionalization from partitions to arbitrary collections of events. Furthermore, the postulate of Jeffrey's probability kinematics, which is rooted in Ramsey's subjectivism, turns out to be a consequence in our frequentist semantics.

I think the book appeals to researchers that are involved in any kind of knowledge processing systems. F.P. conditionalization is a straightforward, fundamental concept that fits our intuition. Furthermore, it creates a clear link from the Kolmogorov system of probability to one of the important Bayesian frameworks. This way, I think it is interesting for anybody who investigates semantics of reasoning systems. The list of these mutually overlapping theories, methods and tools includes, without preference, multivariate data analysis, Bayesian frameworks, fuzzy logic, many-valued logics, conditional logic, Nilsson probabilistic logic, probabilistic model checking and also current efforts in unifying probability theory and logics such as the current rational programming.

Tallinn, August 2017 *Dirk Draheim*

Contents

Chapter 1
Introduction

This book provides a frequentist semantics for conditionalization on partially known events which is given as a straightforward generalization of classical conditional probability via so-called probability testbeds. For this purpose, we compare it with an operational semantics of classical conditional probability that is made precise in terms of sequences of so-called conditional events and accompanied by a corresponding instance of the strong law of large numbers. We analyze the resulting partial conditionalization, that we call frequentist partial (F.P.) conditionalization, from different angles, i.e., with respect to partitions, segmentation, independence, and chaining. It turns out that F.P. conditionalization generalizes Jeffrey conditionalization from partitions to arbitrary collections of events, this way opening it for reassessment and a range of potential applications. A counterpart of Jeffrey's rule for the case of independence holds in our frequentist semantics. We compare this result to Jeffrey's commutative chaining of independent updates and the corresponding possible worlds' belief function. Furthermore, the postulate of Jeffrey's probability kinematics, which is rooted in Ramsey's subjectivism and which can be shown analytically equivalent to Donkin's principle, turns out to be a consequence in our frequentist semantics. This way, the book bridges between the Kolmogorov system of probability and one of the important Bayesian frameworks. Then, we will see that an alternative preservation result, i.e., for conditional probabilities under all updated events, holds in our frequentist semantics and exploit it to discuss a possible redesign of the axiomatic basis of probability kinematics. Furthermore, the book looks at desirabilities, which are again a central concept in Ramsey's subjectivism and Jeffrey's logic of decision, and proposes a more fine-grained analysis of desirabilities a posteriori.

The book takes probabilistic reasoning as the subject of investigation. In the past decades, we have seen immense interest in probabilistic reasoning techniques, just think of the artificial intelligence and the data mining community. The book aims to build a path of mitigation between the Bayesian world view and the frequentist world view by giving a frequentist semantics to partial conditionalization. Our approach is reductionist. We take a single, important Bayesian notion as our starting point, i.e., Jeffrey conditionalization by Richard C. Jeffrey [79, 81–87, 89, 92].

D. Draheim, *Generalized Jeffrey Conditionalization*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-3-319-69868-7_1

1.1 From Conditional Probability to Partial Conditionalization

We give a frequentist semantics of conditionalization on arbitrary many partially known events. It turns out that in the special case of non-overlapping events our semantics meets Jeffrey conditionalization. It could be said that we achieve two things, i.e., a generalization of Jeffrey conditionalization plus a pure frequentist interpretation of partial conditionalization. To get the point, first consider the classical notion of conditional probability. Given events A and B, we know that the conditional probability $\mathsf{P}(A|B)$ is defined as

$$\boxed{\mathsf{P}(A|B) = \mathsf{P}(AB)/\mathsf{P}(B)} \tag{1.1}$$

The value $\mathsf{P}(A|B)$ is called the conditional probability of A under condition B [95]. Now, what is $\mathsf{P}(A|B)$ intended to mean? One way to understand it is as follows. The event B has actually occurred, i.e., we have actually observed the event B. Now, $\mathsf{P}(A|B)$ expresses the probability that event A has also occurred.

Now, we could say that $\mathsf{P}(A|B)$ expresses the idea that the probability of B changes from an old probability $\mathsf{P}(B)$, which is, in general different to 100%, into a new probability of 100%. Here, the old probabilities $\mathsf{P}(AB)$, $\mathsf{P}(A\overline{B})$, $\mathsf{P}(\overline{A}B)$, $\mathsf{P}(\overline{AB})$ etc. can be called *a priori* probabilities, whereas the new 100%-probability of B and the new $\mathsf{P}(A|B)$-probability of A can be called *a posteriori* probabilities.

Now, why allowing the *a priori* probability of the condition B of a conditional probability $\mathsf{P}(A|B)$ to be changed into a 100% probability only? Why not allowing it to change into an arbitrary new probability b? Allowing this is exactly what a non-classical conditional probability might be about and what we want to call a partial conditionalization in the sequel. Given a list of events $B_1,\ldots,B_m$ and a list of *a posteriori* probabilities $b_1,\ldots,b_m$, we introduce the notion of probability of event A conditional on the *a posteriori* probability specifications $B_1 \equiv b_1,\ldots,B_m \equiv b_m$ and introduce the following notation for it:

$$\mathsf{P}(A\,|\,B_1 \equiv b_1,\ldots,B_m \equiv b_m) \tag{1.2}$$

It is Richard C. Jeffrey who investigates conditional probabilities of the form in Eqn. (1.2) and gives concrete probability values to them, albeit he uses a different notation for them that we will discuss later. He considers those situations, in which the events $B_1,\ldots,B_m$ form a partition of the outcome space. In these cases, Jeffrey gives the following value to partial conditionalizations:

$$\boxed{\mathsf{P}(A\,|\,B_1 \equiv b_1,\ldots,B_m \equiv b_m)_{\mathrm{J}} = \sum_{\substack{i=1\\ \mathsf{P}(B_i)\neq 0}}^{m} b_i \cdot \mathsf{P}(A\,|\,B_i)} \tag{1.3}$$

The semantics for partial conditionalization expressed by Eqn. (1.3) is known as Jeffrey conditionalization and often also called Jeffrey's rule. We have marked the conditionalization in Eqn. (1.3) with a J as index to distinguish it from the our general notion of partial conditionalization in Eqn. (1.2). Actually, we want to exploit