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 nonclassical 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 the notation $\mathsf{P}(A \mid B_1 \equiv b_1,\ldots,B_m B_m \equiv b_m)$ for any kind of partial conditionalization, depending on the context, and so we use it for our own semantics of partial conditionalization, called F.P. conditionalization, in due course.

What can be the intended meaning of a partial update specification $B \equiv b$? Jeffrey explains this with the notion of *degree of belief* and justifies the Eqn. (1.3) by so-called probability kinematics. We take a different approach. We will give a frequentist semantics to $\mathsf{P}(A \mid B_1 \equiv b_1,\ldots,B_m \equiv b_m)$. In a sense, we just generalize the notion of classical conditional probability $\mathsf{P}(A|B)$. Let us have a look at $\mathsf{P}(A|B)$ again. We have said that we might understand it as follows. B has actually occurred and $\mathsf{P}(A|B)$ stands for the probability that A has also occurred. But what can this practically mean? We can think of it as follows. We conduct the original experiment over and over again, i.e., a sufficiently large number of times, and then throw away all the completed experiments in which event B did not occur. Throwing away just means that we ignore them as if they have not occurred. So, we consider only those completed experiments, in which B occurred. Among those, we can expect that event A occurs $\mathsf{P}(A|B)$ times. Why? Intuitively, because $\mathsf{P}(B)$ is the expected relative number of occurrences of B and, similarly, $\mathsf{P}(AB)$ is the expected relative number of occurrences of AB, so that we can expect $\mathsf{P}(AB)/\mathsf{P}(B)$ as the relative number of occurrences of A in those completed experiments that yielded B. This intuition adheres completely to Andrey Kolmogorov's original explication of probability theory in [95], which characterizes a probability value as the frequency of an event in a repeated experiment. Formally, this understanding of the conditional probability $\mathsf{P}(A|B)$ is backed by the laws of large numbers. And it is exactly this understanding of the conditional probability that guides us in giving a frequentist semantics to the partial conditionalization $\mathsf{P}(A \mid B_1 \equiv b_1,\ldots,B_m \equiv b_m)$.

Now, how do we give semantics to $\mathsf{P}(A \mid B_1 \equiv b_1,\ldots,B_m \equiv b_m)$? We consider repeated experiments of such lengths, in which statements of the form $B_i \equiv b_i$ make sense frequentistically, i.e., the probability b_i can be interpreted as the frequency of B_i and can potentially be observed. Then we reduce the notion of partial conditionalization to the notion of classical conditional probability. We consider the expected value of the frequency of A, i.e., the average occurrence of A, conditional on the event that the frequencies of events B_i adhere to the new probabilities b_i. Now, we can speak of the b_is as frequencies. We can speak of pairs $B_i \equiv b_i$ also as frequency specifications or probability specifications. We call the resulting notion frequentist partial conditionalization (F.P.) conditionalization. Now, how do we write down, what we just have said, in standard notation? Given an adequate number n of experiment repetitions, we will define the F.P. conditionalization bounded by n and denoted by $\mathsf{P}^n(A \mid B_1 \equiv b_1,\ldots,B_m \equiv b_m)$ as follows:

$$\boxed{\mathsf{P}^n(A \mid B_1 \equiv b_1,\ldots,B_m \equiv b_m) = \mathsf{E}(\overline{A^n} \mid \overline{B_1{}^n} = b_1,\ldots,\overline{B_m{}^n} = b_m)} \tag{1.4}$$

In general, the F.P. conditionalization $\mathsf{P}(A \mid B_1 \equiv b_1,\ldots,B_m \equiv b_m)$ will be defined via the limit of expected values in all repeated experiments of adequate numbers of repetitions. Then bounded F.P. conditionalizations $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ of the form Eqn. (1.4) are approximations to an F.P. conditionalization. In some important cases, the F.P. conditionalizations $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ are equal for all adequate numbers of repetitions n. First, this is so in case of Jeffrey conditionalization, in which the events B_1 through B_m form a partition of the outcome space. Second, this is so in case the events B_1 through B_m are mutually independent.

Remarks on Notation and Terminology

Jeffrey also gives a value to the probability of an event A conditional on a single probability specification $B \equiv b$ as follows:

$$\mathsf{P}(A \mid B \equiv b)_{\mathrm{J}} = b \cdot \mathsf{P}(A \mid B) + (1-b) \cdot \mathsf{P}(A \mid \overline{B}) \tag{1.5}$$

Equation (1.5) can be considered as an instance of Eqn. (1.3) due to the fact that $\mathsf{P}(A \mid B \equiv b)_{\mathrm{J}}$ can be rewritten as $\mathsf{P}(A \mid B \equiv b, \overline{B} \equiv (1-b))_{\mathrm{J}}$. Equation (1.5) itself is also often called Jeffrey conditionalization or Jeffrey's rule.

The notation for partial conditionalization $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ is an arbitrary choice, in particular, the notation for probability specifications $B \equiv b$. A probability specification $B \equiv b$ expresses the fact that the probability of an event B changes from an *a priori* probability $\mathsf{P}(B)$ to a new, *a posteriori* probability b. So, maybe some notation such as $\mathsf{P}(B) \rightsquigarrow b$ or $\mathsf{P}(B) := b$ would be more appropriate. We have chosen the short notation $B \equiv b$ for the sake of readability.

Note that our notation for Jeffrey conditionalization $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)_{\mathrm{J}}$ in Eqn. (1.3) is oriented towards our own notation for partial conditionalizations. The original notation of Jeffrey in [87] is different. Jeffrey denotes all *a priori* probabilities as $prob(A)$ and all *a posteriori* probabilities as $PROB(A)$ so that Eqn. (1.3) looks like

$$PROB(A) = \sum_{\substack{i=1 \\ \mathsf{P}(B_i) \neq 0}}^{m} PROB(B_i) \cdot prob(A \mid B_i) \tag{1.6}$$

In particular, Jeffrey does not make explicit the conditions of $PROB(A)$ as we do with the $B_i \equiv b_i$ in $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$. Rather, Eqn. (1.6) relies on the context making clear under which conditions $PROB(A)$ is a conditionalization. Actually, the question of explicit or implicit notation has no semantic implications. With respect to Jeffrey conditionalization and other kinds of possible world conditionalizations, like the one considered by William F. Donkin, this question is a merely notational issue. Actually, in early literature Jeffrey uses yet another option halfway between the implicit and the explicit notations, with explicit update values $b_1, \ldots, b_m$ that are maintained in the context. For F.P. conditionalization, the explicit notation is essential. We define the conditionalization, we do not simply postulate it. Therefore, we need to proof first that the *a posteriori* probability $\mathsf{P}(B_i|B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ of an event B_i equals the value b_i that we assign to it. In possible world frameworks, this is not a question; it is just taken per establishment of the conditionalization. Anyhow, $\mathsf{P}(B_i|B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ is the first thing that we will proof for F.P.

semantics and then could use an implicit notation, at least in large parts, also for F.P. conditionalization. Again, the question of explicit vs. implicit notation does not matter conceptually, in particular, when it comes to comparisons between F.P. semantics and Jeffrey conditionalization or other possible world semantics.

When we deal with F.P. conditionalization, we only use the explicit notation. The explicit notation shows a neat correspondence between the frequency specifications $B \equiv b$ and $\overline{B^n} = b$. Also, the explicit notation is easier to maintain in proofs and argumentations. When we deal with Jeffrey conditionalization or other conditionalizations from the Bayesian realm as, e.g., found in Donkin's principle later, we feel free to use either the implicit or the explicit notation. Usually, we want to switch to the explicit notation when it comes to comparisons between different conditionalization frameworks.

1.2 Background, Motivation, Perspectives and Outlook

We make conditionalization over partial update specifications the subject of our interest. We do not want to take a position in favor or against one of the many important logical apparatuses, reasoning frameworks. We even try to avoid statements in favor or against one of the big schools, the Bayesian or the frequentist. Rather, we want to provide a bridge between the Bayesian worldview and the frequentist world. F.P. conditionalization has the following key characteristics. F.P. conditionalization semantics if consistent with the original explication of probability theory as characterized by Andrey Kolmogorov. It makes no use of further concepts from psychology, epistemology, decision theory etc. F.P. conditionalization turns out to be consistent with Jeffrey conditionalization, i.e., we can prove that it yields the same values as Jeffrey conditionalization – in the special case that is treated by Jeffrey conditionalization, i.e., partitions. F.P. conditionalization is a generalization of Jeffrey conditionalization. Jeffrey conditionalization is defined for partitions only. F.P. conditionalization is different. It is defined, declaratively, for arbitrary lists of events. Jeffrey conditionalization relies on the postulate that an *a priori* conditional probability with respect to one of the updated events remains unchanged after conditionalization, i.e., after update of the condition event. The justification of Jeffrey's rule needs this so-called kinematics postulate. This is different in the F.P. framework. Here, Jeffrey's rule is a consequence of the definition of F.P. semantics. F.P. conditionalization creates a link from the Kolmogorov system of probability to one of the important Bayesian frameworks, i.e., Jeffrey's logic of decision.

The approach of this book is formal. We take a notion from the Bayesian realm, i.e., conditionalization with respect to partial updates, and will equip it with a formal, rigor semantics from the frequentist realm. We hope that the results are useful for the reader for his or her own reception, in his or her own, individual assessment of reasoning techniques and approaches.

Frequentism can be clearly identified with what Julian Jaynes calls the Kolmogorov system of probability [78]. Here, we are stable and safe, I mean from a technical viewpoint. The frequentist approach is the perspective of Bernoulli and his Golden Theorem. It has an authoritative explication, the observation of a repeated experiment in a long series of repetitions [94–96]. Once formalized, e.g., in today's standard formalization in measure theory, this explication is reinforced or let's say confirmed by the laws of large numbers and the central limits theorems that follow from this formalization. This way, the frequentist viewpoint emerged into a particularly consistent and strong argumentative framework. So, when it comes to frequentism, we have a very clear understanding of what this is about. Here, the question of objectivism is not relevant for the developments in the book, and we rather would like to refer to the discussions conducted by Jerzy Neyman [116, 118] with respect to this. We use frequentism rather as synonym for the Kolmogorov theory together with its standard explication. Once we have agreed that this is not an oversimplification, it is fair to say, again, that we have a very clear understanding of what frequentism is about.

When it comes to Bayesianism, things are different. There is no such single, closed apparatus as with frequentism. Instead, as you know, there is a great variety of important approaches and methodologies, with different flavors in objectives and explications [64, 153, 155]. Think of Bruno de Finetti [58, 59] with his Dutch book argument and Frank P. Ramsey [127, 129] with his representation theorem [128]. Think of Julian Jaynes [76], who starts from improving statistical reasoning with his application of maximal entropy [77], and from there transcends into an agent-oriented explanation of probability theory [78]. Also, think of Judea Pearl [123], who eventually transcends probabilistic reasoning by systematically incorporating causality into his considerations [124, 125]. Bayesian approaches have in common that they rely, at least in crucial parts, on notions other than frequencies to explain probabilities, among the most typical are degrees of belief, degrees of preference, degrees of plausibility, degrees of validity, degrees of confirmation or, on the opposite side, degrees of uncertainty. Also, they have in common, that they, more or less, stress and exploit notions of probability update, typically traced back to Bayes' rule and then usually called Bayesian update.

Let us have a look at Rudolf Carnap's Probability–1 and Probability–2 [23, 24] that he uses to distinguish between two fundamentally different concepts of probability. It is Probability–2 that refers the frequentist interpretation of probability theory. Now, Carnap's Probability–1 is clearly about degree of confirmation. But there are more characteristics that distinguish Probability–1 from Probability–2 and it is interesting to have a look at them. For Carnap, Probability–2 is clearly an empirical concept; it is about observation of events. And sentences in Probability–2 are sentences about factual states of affairs. And now, with Probability–1 sentences are not about observations, there are considered purely analytical entities. He uses terms such as semantical and logical to characterize this notion of analytical probability description.

All the single important Bayesian frameworks are each highly sophisticated and elaborated. However, as we said, Bayesianism does not stand for a single closed framework, but rather for a variety of frameworks that share, more or less, some common characteristics. In this book, we bridge between frequentism and one of the important Bayesian frameworks, i.e., the logic of decision of Richard C. Jeffrey. First, we give our frequentist semantics to a general notion of partial conditionalization. Then it turns out that our partial conditionalization semantically meets Jeffrey conditionalization, which is the core concept of Jeffrey's probability kinematics, which is again a or even *the* crucial building block of Jeffrey's logic of decision. It turns out that the found concept is, in a sense, is a true generalization of Jeffrey conditionalization, as Jeffrey conditionalization is defined only for partitions, i.e., collections of condition events that form a partition. As opposed to that, the found partial conditionalization works for arbitrary, also overlapping, condition events.

In probability kinematics , a certain property of probabilities and expectations is assumed as given. It is postulated. It is the invariance of a conditional probability of a target event over one of the condition events *a posteriori*, i.e., after conditionalization. The postulate is crucial for probability kinematics for two reasons. First, it is needed to justify the definition of Jeffrey conditionalization. Actually, the definition of Jeffrey conditionalization follows from the postulate via the law of total probability. Second, and equally important, it brings Jeffrey's logic of decision together, conceptually, with the important Bayesian framework of Frank P. Ramsey. Now, our frequentist conditionalization sheds some new light on the postulate of probability kinematics. In our frequentist semantics the postulate follows from the definition of our conditionalization, whereas, in the logic of decision, it is assumed as a basic concept, as the axiomatic basis so to speak. Furthermore, it turns out, that the special case of partitions is essential to probability kinematics, as Jeffrey conditionalization can only be derived from its postulate and therefore justified if the events are guaranteed to form a partition.

When I first dealt with F.P. conditionalization, I was aware of Jeffrey conditionalization; however, my efforts were not particularly intended towards Jeffrey conditionalization. The idea was to generalize classical conditional probabilities, so that partially known events make sense frequentistically. The idea was then to look at sequences of base experiments in which partial probability specifications can be observed and this way make sense of partially known events. Now, the question was what we can expect as the relative occurrence of a target event in such probability testbeds. When I first dealt this notion, I was immediately thinking about it in terms of its combinatorial solution, see Sect. 4.1.2, i.e., a true frequentist approach so to speak. The combinatorial solution of the involved multivariate Bernoulli distributions is a straightforward generalization of the combinatorial solution for basic Bernoulli distributions. Declaratively, the searched concept is the conditional expected value over partial update specifications, and this is how our notion of partial conditionalization is defined.

In this book we are interested in probabilistic reasoning in its own right. In [44] we are interested in reasoning about systems that contain programmable components and act in a probabilistic environment. We take a reductionist approach in [44] and investigate the probabilistic typed lambda-calculus, which is a minimal functional programming language plus a probabilistic programming choice construct. The work in [44] is motivated by previous work on the engineering and operation of business process technology [5, 40, 42, 48–50]. Given the relevance of probabilistic reasoning to business processes, compare with [25, 29, 30], this closes the loop to the current book. In [43] we deal with reflective constraint writing, which is the counterpart of full reflective programming with respect to constraint writing. Reflective constraint writing has its application in multi-level systems engineering and maitenance [4, 11, 41, 45–47, 51, 52, 108], however, its original motivation was in semantic clarification of multi-level modelling [53, 54, 70]. With the investigated decoupling of the intention vs. the pragmatics of multi-level modeling languages the work touches the considerations in the current book.

1.3 Chapter Overview

In Chap. 2 we define F.P. conditionalization. We start the chapter with a recap on how to model repeated experiments. Then we define F.P. conditionalization as the expected value of an event in a testbed that adheres to frequency constraints. It also presents, as a Lemma, an alternative definition of F.P. conditionalization as a conditional probability with respect to adequate testbeds. The second definition is easier to handle in proofs and argumentations and is, therefore, the *de facto* definition of F.P. conditionalization on many occasions. The chapter proceeds with looking into a particularly intuitive kind of F.P. conditionalizations, i.e., such that are projective in the target event. Next, the chapter looks into an important technique for transforming F.P. testbed specifications, i.e., shortening a testbed while adjusting the frequencies of the involved frequency specifications. Finally, we look into an operational semantics of classical conditional probabilities resulting into the notion of so-called conditional event and the corresponding instance of the strong law of large numbers.

Chapter 3 deals with classical Jeffrey conditionalization. Jeffrey conditionalization deals with the special case of condition events that form a partition. We show that F.P. conditionalization meets Jeffrey conditionalization. We start with the basic case of a single condition event and proceed with Jeffrey conditionalization in general. With Chap. 3 Jeffrey conditionalization is technically embedded into F.P. conditionalization.

Chapter 4 investigates properties of full, general F.P. conditionalizations. We look into how F.P. conditionalization can be computed. We do so by giving both a recursive definition as well as a combinatorial characterization. Next, we present a segmentation Lemma which draws a further, convenient analog to the special case of Jeffrey conditionalization. Next, we will see that independence of condition events is preserved by F.P. conditionalization. As a consequence, it is easy to determine the value of each F.P. conditionalization in case of independent conditions. The independence result is important as it again reinforces our intuition about F.P. conditionalization. Next, we will see in how far classical conditional probabilities *a priori* are preserved under F.P. update. Furthermore, we investigate the behavior of expected values under after F.P. conditionalization.

In Chap. 5 we conduct a discussion of Jeffrey's probability kinematics and F.P. semantics. We discuss Jeffrey's postulate and compare Jeffrey's commutative chaining of independent events with F.P. segmentation of independent events. Against this background, we discuss the redesign of the axiomatic basis of probability kinematics. Furthermore, we will discuss the subjectivist concept of desirabilities and propose a fine-grained investigation of desirabilities after partial update. Furthermore, we will investigate the correspondence between Donkin's principle and Jeffrey's postulate. It turns out that Donkin's principle and Jeffrey's postulate are equivalent.

Appendix A provides bibliographic notes on related work with respect to Richard C. Jeffrey's writings, contingency tables for given observations and marginals, rationales of probability kinematics, closeness of probability measures, non-commutativity of Jeffrey conditionalization, Dempster-Shafer theory, the maximal entropy principle, Jeffrey desirabilities and a series of further relevant topics. Appendix B serves as a repository for some mathematical definitions, whereas Appendix C compiles some technical Lemmas that are important in proofs and argumentations of the book.

http://www.springer.com/978-3-319-69867-0

Generalized Jeffrey Conditionalization
A Frequentist Semantics of Partial Conditionalization
Draheim, D.
2017, X, 106 p., Softcover
ISBN: 978-3-319-69867-0