Freely Combining Partial Knowledge in Multiple Dimensions
(Extended Abstract)

Dirk Draheim(✉)

Large-Scale Systems Group, Tallinn University of Technology,
Akadeemia tee 15a, 12618 Tallinn, Estonia
dirk.draheim@ttu.ee

Abstract. F.P. conditionalization (frequentist partial conditionalization) allows for combining partial knowledge in arbitrary many dimensions and without any restrictions on events such as independence or partitioning. In this talk, we provide a primer to F.P. conditionalization and its most important results. As an example, we proof that Jeffrey conditionalization is an instance of F.P. conditionalization for the special case that events form a partition. Also, we discuss the *logics* and the *data science* perspective on the matter.

Keywords: F.P. conditionalization · Jeffrey conditionalization
Data science · Statistics · Contingency tables · Reasoning systems
SPSS · SAS · R · Phyton/Anaconda · Cognos · Tableau

1 A Primer on F.P. Conditionalization

In [1] we have introduced F.P.conditionalization (frequentist partial conditionalization), which allows for conditionalization on partially known events. An F.P. conditionalization $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ is the probability of an event A that is conditional on a list of event-probability specifications $B_1 \equiv b_1$ through $B_m \equiv b_m$. A specification pair $B \equiv b$[1][2] stands for the assumption that the probability of B has somehow changed from a previously given, *a priori* probability $\mathsf{P}(B)$ into a new, *a posteriori* probability b. Consequently, we expect that $\mathsf{P}(B \mid B \equiv b) = b$ as well as $\mathsf{P}(A \mid B \equiv \mathsf{P}(B)) = \mathsf{P}(A)$. Similarly, we expect that classical conditional probability becomes a special case of F.P. conditionalization, i.e., that $\mathsf{P}(A|B_1 \cdots B_m)$ equals $\mathsf{P}(A \mid B_1 \equiv 100\%, \ldots, B_m \equiv 100\%)$ and, similarly, $\mathsf{P}(A|\overline{B_1} \cdots \overline{B_m})$ equals $\mathsf{P}(A \mid B_1 \equiv 0\%, \ldots, B_m \equiv 0\%)$.

But what is the value of $\mathsf{P}(A|B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ in general? We have given a formal, frequentist semantics to it. We think of conditionalization as taking

[1] Alternative notations for $B \equiv b$ such as $\mathsf{P}(B) \rightsquigarrow b$ or $\mathsf{P}(B) := b$ might be considered more intuitive. We have chosen the concrete notation $B \equiv b$ for the sake of brevity and readability.

[2] We also use $\mathsf{P}_{B_1 \equiv b_1, \ldots, B_m \equiv b_m}(A)$ as notation for $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$.

© Springer Nature Switzerland AG 2018
T. K. Dang et al. (Eds.): FDSE 2018, LNCS 11251, pp. 3–11, 2018.
https://doi.org/10.1007/978-3-030-03192-3_1

place in chains of repeated experiments, so-called probability testbeds, of sufficient lengths. As a first step, we introduce the notion of F.P. conditionalization bounded by n which is denoted by $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$. We consider repeated experiments of such lengths n, 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, i.e., classical conditional expected value to be more precise. 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. Next, we define (general/unbounded) F.P. conditionalization by bounded F.P. conditionalization in the limit.

Definition 1 (Bounded F.P. Conditionalization). Given an i.i.d.sequence (independent and identically distributed sequence) of multivariate characteristic random variables $(\langle A, B_1,..., B_m\rangle_{(j)})_{j\in\mathbb{N}}$, a list of rational numbers $b_1,..., b_m$ and a bound $n \in \mathbb{N}$ such that $0 \leqslant b_i \leqslant 1$ and $nb_i \in \mathbb{N}$ for all b_i in $b_1,..., b_m$. We define the *probability of A conditional on $B_1 \equiv b_1$ through $B_m \equiv b_m$ bounded by n*, which is denoted by $\mathsf{P}^n(A \mid B_1 \equiv b_1,..., B_m \equiv b_m)$, as follows:

$$\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) = \mathsf{E}(\overline{A^n} \mid \overline{B_1}^n = b_1, \ldots, \overline{B_m}^n = b_m) \tag{1}$$

Definition 2 (F.P. Conditionalization). Given an i.i.d.sequence of multivariate characteristic random variables $(\langle A, B_1,..., B_m\rangle_{(j)})_{j\in\mathbb{N}}$ and a list of rational numbers $b = b_1,..., b_m$ such that $0 \leqslant b_i \leqslant 1$ for all b_i in b and $lcd(b)$ denotes the smallest $n \in \mathbb{N}$ such that $nb_i \in \mathbb{N}$ for all b_i in $b = b_1,..., b_m$.[3] We define the *probability of A conditional on $B_1 \equiv b_1$ through $B_m \equiv b_m$*, denoted by $\mathsf{P}(A \mid B_1 \equiv b_1,..., B_m \equiv b_m)$, as follows:

$$\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) = \lim_{k\to\infty} \mathsf{P}^{\,k\cdot lcd(b)}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) \tag{2}$$

As a first result, we observe that bounded F.P. conditionalization can be expressed more compact, without conditional expectation, merely in terms of conditional probability, i.e., we have that the following holds for any bounded F.P. conditionalization:

$$\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) = \mathsf{P}(A \mid \overline{B_1}^n = b_1, \ldots, \overline{B_m}^n = b_m) \tag{3}$$

In most proofs and argumentations we use the more convenient form in Eq. (3) instead of the more intuitive form in Definition 1.

In general, an F.P. conditionalization $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ is different from all of its finite approximations of the form $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$. In some interesting special cases, we have that the F.P. conditionalizations are equal to all of their finite approximations; i.e., it is the case if the condition events $B_1 \equiv b_1$ through $B_m \equiv b_m$ are independent or if the condition events form a partition.

[3] $lcd(b)$ is the *least common denominator* of $b = b_1,..., b_m$.

The case in which the condition events form a partition is particularly interesting. This is so, because this case makes Jeffrey conditionalization [2–4], value-wise, an instance of F.P. conditionalization as we will discuss further in Sect. 2. In case the conditions events $B_1 \equiv b_1$ through $B_m \equiv b_m$ form a partition, we have that the value of $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ is a weighted sum of conditional probabilities $b_i \cdot \mathsf{P}(A|B_i)$, compare with Eq. (5). This is somehow neat and intuitive. Take the simple case of an F.P. conditionalization $\mathsf{P}(A|B \equiv b)$ over a single event B. Such an F.P. conditionalization can be represented differently as an F.P. conditionalization over two partioning events $B_1 = B$ and $B_2 = \overline{B}$, i.e., $\mathsf{P}(A \mid B \equiv b\,,\ \overline{B} \equiv 1-b)$. Therefore we have that

$$\mathsf{P}(A|B \equiv b) = b \cdot \mathsf{P}(A|B) + (1-b) \cdot \mathsf{P}(A|\overline{B}) \tag{4}$$

Equation 4 is highly intuitive: it feels natural that the direct conditional probability $P(A|B)$ should be somehow (proportionally) lowered by the new probability b of event B, similarly, we should not forget that the event $\overline{B}$ can also appear, i.e., with probability $1-b$ and should also influence the final value – symmetrically. So, the b-weighted average of $P(A|B)$ and $P(A|\overline{B})$ as expressed by Eq. (4) seems to be an educated guess. Fortunately, we do not need such an appeal to intuition. In our framework, Eqs. (4) and (5) can be proven correct, as a consequence of probability theory.

Theorem 3 (F.P. Conditionalization over Partitions). *Given an F.P. conditionalization* $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ *such that the events* $B_1, \ldots, B_m$ *form a partition, and, furthermore, the frequencies* $b_1, \ldots, b_m$ *sum up to one, we have the following:*

$$\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) = \sum_{\substack{1 \leqslant i \leqslant m \\ \mathsf{P}(B_i) \neq 0}} b_i \cdot \mathsf{P}(A \mid B_i) \tag{5}$$

Proof. See [1].

Table 1 summarizes interesting properties of F.P. conditionalization. Proofs of all properties are provided in [1]. Property (a) is a basic fact that we mentioned earlier; i.e., an updated event actually has the probability value that it is updated to. Properties (b) and (c) deal with condition events that form a partition and we have treated them with Theorem 3. Properties (d) and (e) provide programs for probabilities of frequency specifications of the general form $\mathsf{P}(\cap_{i \in I} B_i^n = k_i)$. Having programs for such probabilities is sufficient to compute any F.P. conditionalization. The equation in (d) is called one-step decomposition in [1] and can be read immediately as a recursive programme specification; compare also with the primer on inductive definitions in [5]. Equation (e) provides a combinatorial solution for $\mathsf{P}(\cap_{i \in I} B_i^n = k_i)$. Equation (e) generalizes the known solution for bivariate Bernoulli distributions [6–8] to the general case of multivariate Bernoulli distributions. Property (f) is called conditional segmentation in [1]. Conditional segmentation shows how F.P. conditionalization

Table 1. Properties of F.P. conditionalization. Values of various F.P. conditionalizations $\mathsf{P}_{\boldsymbol{B}}(A) = \mathsf{P}(A|B_1 \equiv b_1,..., B_m \equiv b_m)$ with frequency specifications of the form $\boldsymbol{B} = B_1 \equiv b_1,..., B_m \equiv b_m$ and condition indices $I = \{1, \ldots, m\}$; probability values (d) and (e) of frequency specifications of the form $\mathsf{P}(\cap_{i\in I} B_i^n = k_i)$. Proofs of all properties are provided in [1].

	Constraint	F.P. Conditionalization
(a)	b_i belongs to $\boldsymbol{B}$	$\mathsf{P}_{\boldsymbol{B}}(B_i) = b_i$
(b)	$m = 1,\ \boldsymbol{B} = (B \equiv b)$	$\mathsf{P}_{\boldsymbol{B}}(A) = b \cdot \mathsf{P}(A\|B) + (1-b) \cdot \mathsf{P}(A\|\overline{B})$
(c)	$B_1,..., B_m$ form a partition	$\mathsf{P}_{\boldsymbol{B}}(A) = \sum_{i=1}^{m} b_i \cdot \mathsf{P}(A \mid B_i)$
(d)	For arbitrary bound n	$\mathsf{P}(\bigcap_{i\in I} B_i^n = k_i) = \sum_{\substack{I' \subseteq I \\ \forall i \in I' . k_i \neq 0 \\ \forall i \notin I' . k_i \neq n}} \mathsf{P}(\bigcap_{i\in I'} B_i, \bigcap_{i\notin I'} \overline{B_i}) \cdot \mathsf{P}(\bigcap_{i\in I'} B_i^{n-1} = k_i - 1, \bigcap_{i\notin I'} B_i^{n-1} = k_i)$
(e)	For arbitrary bound n	$\mathsf{P}(\bigcap_{i\in I} B_i^n = k_i) = \sum \left(\frac{n!}{\prod_{I'\subseteq I} \rho(I')!} \times \prod_{I'\subseteq I} \mathsf{P}(\bigcap_{i\in I'} B_i, \bigcap_{i\notin I'} \overline{B_i})^{\rho(I')} \right)$ $\rho : \mathbb{P}(I) \rightarrow \mathbb{N}_0$ $\forall i \in I . k_i = \sum\{\rho(I') \mid I' \subseteq I \wedge B_i \in I'\}$ $n = \sum\{\rho(I') \mid I' \subseteq I\}$
(f)	–	$\mathsf{P}_{\boldsymbol{B}}(A) = \sum_{\substack{(\zeta_i \in \{B_i, \overline{B_i}\})_{i\in I} \\ \mathsf{P}(\bigcap_{i\in I} \zeta_i) \neq 0}} \mathsf{P}(A\| \bigcap_{i\in I} \zeta_i) \cdot \mathsf{P}(\bigcap_{i\in I} \zeta_i \| \bigcap_{i\in I} B_i \equiv b_i)$
(g)	$B_1,..., B_m$ are independent	$\mathsf{P}_{\boldsymbol{B}}(B_1,..., B_k) = b_1 b_2 \cdots b_k$
(h)	$B_1,..., B_m$ are independent	$\mathsf{P}_{\boldsymbol{B}}(B_1,..., B_m) = \mathsf{P}_{\boldsymbol{B}}(B_1) \cdots \mathsf{P}_{\boldsymbol{B}}(B_m)$
(i)	$B_1,..., B_m$ are independent	$\mathsf{P}_{\boldsymbol{B}}(A) = \sum_{\substack{I' \subseteq I \\ \mathsf{P}(\bigcap_{i\in I'} B_i, \bigcap_{i\notin I'} \overline{B_i}) \neq 0}} \left(\mathsf{P}(A\| \bigcap_{i\in I'} B_i, \bigcap_{i\notin I'} \overline{B_i}) \cdot \prod_{i\in I'} b_i \cdot \prod_{i\notin I'} (1-b_i) \right)$
(j)	A is independent of $B_1,..., B_m$	$\mathsf{P}_{\boldsymbol{B}}(A) = \mathsf{P}(A)$
(k)	$B_1 \equiv 100\%,..., B_i \equiv 100\%$ $B_{i+1} \equiv 0\%,..., B_m \equiv 0\%$	$\mathsf{P}_{\boldsymbol{B}}(A) = \mathsf{P}(A\|B_1,..., B_i, \overline{B_{i+1}},..., \overline{B_m})$
(l)	$B_1,..., B_m$ form a partition *or* $B_1,..., B_m$ are independent $B_1 \equiv \mathsf{P}(B_1),..., B_m \equiv \mathsf{P}(B_m)$	$\mathsf{P}_{\boldsymbol{B}}(A) = \mathsf{P}(A)$
(m)	$B_1,..., B_m$ form a partition	$\mathsf{P}_{\boldsymbol{B}}(AB_i) = b_i \cdot \mathsf{P}(A\|B_i)$
(n)	$B_1,..., B_m$ form a partition	$\mathsf{P}_{\boldsymbol{B}}(A\|B_i) = \mathsf{P}(A\|B_i)$
(o)	$B_1,..., B_m$ are independent	$\mathsf{P}_{\boldsymbol{B}}(A, B_1,..., B_m) = b_1 \cdots b_m \cdot \mathsf{P}(A\|B_1,..., B_m)$
(p)	–	$\mathsf{P}_{\boldsymbol{B}}(A\|B_1,..., B_m) = \mathsf{P}(A\|B_1,..., B_m)$

generalizes Jeffrey conditionalization by dropping the partitioning constraint on events. Conditional segmentation is also often useful as helper Lemma. Properties (g) and (h) are important; they reveal how F.P. conditionalization behaves in case of independent condition events. Property (i) deals with the case that a target event is independent of the condition events. Property (k) has been mentioned earlier; it is about how F.P. conditionalization meets classical conditional probability. Property (l) generalizes the basic fact that $\mathsf{P}(A \mid B \equiv \mathsf{P}(B)) = \mathsf{P}(A)$ to lists of condition events. Properties (m) through (p) all deal with cases, in which condition events also appear, in some way, in the target event. Properties (m) through (p) are highly relevant in the discussion of Jeffrey's probability kinematics and other Bayesian frameworks with possible-world semantics. Actually, property (n) is an F.P. version of what we call Jeffrey's postulate.

Table 2. Properties of F.P. conditional expectations. Values of various F.P. expectations $\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu \mid A)$, with frequency specifications $\boldsymbol{B} = B_1 \equiv b_1, \ldots, B_m \equiv b_m$ and condition indices $I = \{1, \ldots, m\}$. Proofs of all properties are provided in [1].

	Constraint	F.P. Expectation
(A)	$B_1, \ldots, B_m$ form a partition	$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu \mid B_i) = \mathsf{E}(\nu \mid B_i)$
(B)	$m = 1,\ \boldsymbol{B} = (B \equiv b)$	$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu \mid A) = \frac{b \cdot \mathsf{P}(A \mid B)\mathsf{E}(\nu \mid AB) + (1-b) \cdot \mathsf{P}(A \mid \overline{B_1})\mathsf{E}(\nu \mid A\overline{B_1})}{b \cdot \mathsf{P}(A \mid B) + (1-b) \cdot \mathsf{P}(A \mid \overline{B})}$
(C)	$B_1,\ldots, B_m$ form a partition	$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu \mid A) = \frac{\sum_{i=1}^{m} b_i \cdot \mathsf{P}(A \mid B_i) \cdot \mathsf{E}(\nu \mid AB_i)}{\sum_{i=1}^{m} b_i \cdot \mathsf{P}(A \mid B_i)}$
(M)	$B_1,\ldots, B_m$ form a partition	$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu \mid AB_i) = \mathsf{E}(\nu \mid AB_i)$
(N)	$B_1,\ldots, B_m$ form a partition	$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}(\_\mid B_i)}(\nu \mid A) = \mathsf{E}(\nu \mid AB_i)$
(O)	$B_1,\ldots, B_m$ are independent	$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu \mid AB_1 \cdots B_m) = \mathsf{E}(\nu \mid AB_1 \cdots B_m)$
(P)	$B_1,\ldots, B_m$ are independent	$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}(\_\mid B_1 \cdots B_m)}(\nu \mid A) = \mathsf{E}(\nu \mid AB_1 \cdots B_m)$

With Table 2 we step from F.P. conditionalization to F.P. conditional expected values, that we also call F.P. conditional expectations or just F.P. expectations for short. Given frequency specifications $\boldsymbol{B} = B_1 \equiv k_1,\ldots, B_m \equiv k_m$, we say that $\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu \mid A)$ is an F.P. expectation. Here, the event A plays the role of the target event; whereas we consider the random variable ν as rather fixed. This way, each property in Table 1 has a corresponding property in terms of F.P. expectations. Table 2 shows some of them[4]. We do not need an own definition for F.P. expectations. We have that $\mathsf{P}_{\boldsymbol{B}}$ is a probability function, so that the corresponding expected values and conditional expected values[5] are defined and we have that

$$\mathsf{E}_{\mathsf{P}_{\boldsymbol{B}}}(\nu : \Omega \longrightarrow D \mid A) = \sum_{d \in D} d \cdot \mathsf{P}_{\boldsymbol{B}}(\nu = d, A) \Big/ \mathsf{P}_{\boldsymbol{B}}(A) \quad (6)$$

In Ramsey's subjectivism [9–11] and Jeffrey's logic of decision [4,12] the notion of *desirability* is a crucial concept. Here, the desirability *des* A of an event A is the conditional expected value of an implicitly given utility ν under the condition A, which also explains why F.P. expectations are an important concept.

2 The Logics Perspective

In his *logic of decision* [13], also called *probability kinematics* [13,14], Richard C. Jeffrey establishes Jeffrey conditionalization. Probabilities are interpreted as

[4] Rows with same letters in Tables 1 and 2 correspond to each other.
[5] The notation E_{P} makes explicit that E belongs to the probability space $(\Omega, \Sigma, \mathsf{P})$.

degrees of believe and the semantics of a probability update is explained directly in terms of a *possible world semantics*. Jeffrey denotes *a priori* probability values as $prob(A)$ and *a posteriori* probability values as $PROB(A)$ and maintains the list of updated events $B_1,...,B_m$ in the context of probability statements[6]. It is assumed that in both the worlds, i.e., the *a priori* and the *a posteriori* world, the laws of probability hold. The probability functions $PROB$ and $prob$ are related by a postulate. The postulate deals exclusively with situations, in which the updated events $B_1,...,B_m$ form a partition. Then, it states that conditional probabilities with respect to one of the updated events are preserved, i.e., we can assume that $PROB(A|B_i) = prob(A|B_i)$ holds for all events A and all events B_i from $B_1,...,B_m$ – just as longs as $B_1,...,B_m$ form a partition. Persi Diaconis and Sandy Zabell call this postulate the J-condition [15,16]. Richard Bradley talks about conservative belief changes [17,18]. We call this postulate the probability kinematics postulate, or also just Jeffrey's postulate for short. We say that Jeffrey's postulate is a bridging statement, as it bridges between the *a priori* world and the *a posteriori* world. Next, Jeffrey exploits this postulate to derive Jeffrey conditionalization, also called Jeffrey's rule, compare with Eq. (5). It is crucial to understand, that the F.P. equivalent of Jeffrey's postulate, i.e., $\mathsf{P}_{\boldsymbol{B}}(A|B_i) = \mathsf{P}(A|B_i)$[7] does not need to be postulated in the F.P. framework, but is a property that simply holds; i.e., it can be proven from the underlying frequentist semantics.

We have seen that F.P. conditionalization creates a clear link from the Kolmogorov system of probability to one of the important Bayesian frameworks, i.e., Jeffrey's logic of decision. When it comes to Bayesianism, there is no such single, closed apparatus as with frequentism [19–23]. Instead, there is a great variety of important approaches and methodologies, with different flavors in objectives and explications [24–26]. We have de Finetti [27,28] with his Dutch book argument and Ramsey [9,11] with his representation theorem [10]. Think of Jaynes [29], who starts from improving statistical reasoning with his application of maximal entropy [30], and from there transcends into an agent-oriented explanation of probability theory [31]. Also, think of Pearl [32], who eventually transcends probabilistic reasoning by systematically incorporating causality into his considerations [33,34]. 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 or degrees of confirmation.

3 The Data Science Perspective

The *data science* perspective is the F.P. perspective *per se*. Current data science has a clear statistical foundation; in practice, we see that data science is

[6] Please note, that the notational differences between between Jeffrey conditionalization and F.P. conditionalization are a minor issue and must not be confused with semantical differences – see [1] for a thorough discussion.

[7] With $\boldsymbol{B} = B_1 \equiv PROB(B_1),...,B_m \equiv PROB(B_m)$.

boosted by statistical packages and tools, ranging from SPSS, SAS over R to Phyton/Anaconda. In practice, the more interactive, multivariate data analytics (as represented by business intelligence tools such as Cognos or Tableau) is still equally important in data science initiatives. Again, the findings of F.P. conditionalization are fully in line with the foundations of multivariate data analytics.

An important dual problem to partial conditionalization is about determining the most likely probability distribution with known marginals for a complete set of observations. This problem is treated by Deming and Stephan in [35] and Ireland and Kullback in [36]. Given two partitions of events $B_1,...,B_s$ and $C_1,...,C_t$, numbers of observations n_{ij} for all possible B_iC_j in a sample of size n and marginals $p_{i\star}$ for each B_i in and $p_{\star j}$ for each C_j, it is the intention to find a probability distribution P that adheres to the specified marginals, i.e., such that $\mathsf{P}(B_i) = p_{i\star}$ for all B_i and $\mathsf{P}(C_j) = p_{\star j}$ for all C_j, and furthermore maximizes the probability of the specified joint observation, i.e., that maximizes the following multinomial distribution[8]:

$$\mathfrak{M}_{n,\,\mathsf{P}(B_1C_1),..,\mathsf{P}(B_1C_t)\,,...,\,\mathsf{P}(B_sC_1),..,\mathsf{P}(B_sC_t)}(n_{11},...,n_{1t},\ldots,n_{s1},...,n_{st})$$

Note that the collection of $s \times t$ events B_sB_t form a partition. The observed values n_{ij} are said to be organized in a two-dimensional $s \times t$ *contingency table*. The restriction to two-dimensional contingency tables is without loss of generality, i.e., the results of [35] and [36] can be generalized to multi-dimensional tables. In comparisons with partial conditionalizations, we treat two events B and C as a 2×2 contingency table with partitions $B_1 = B$, $B_2 = \overline{B}$, $C_1 = C$ and $C_2 = \overline{C}$. Now, [35] approaches the optimization by least-square[9] adjustment, i.e., by considering the probability function P that minimizes χ^2, whereas [36] approaches the optimization by considering the probability function P that minimizes the Kullback-Leibler number $I(\mathsf{P},\mathsf{P}')$[10] with $\mathsf{P}'(B_iC_j) = n_{ij}/n$; compare also with [37,38]. Both [35,39] and [36] use iterative procedures that generates BAN (best approximatively normal) estimators for convergent computations of the considered minima; compare also with [40,41].

4 Conclusion

Statistics is the language of science; however, the semantics of probabilistic reasoning is still a matter of discourse. F.P. conditionalization provides a frequentist semantics for conditionalization on partially known events. It 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. 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.

8 $\mathfrak{M}_{n,p_1,\ldots,p_m}(k_1,\ldots,k_m) = (n!/(k_1!\cdots k_m!)) \cdot p_1^{k_1}\cdots p_m^{k_m}$.

9 $\chi^2 = \sum_{i=1}^{s}\sum_{j=1}^{t}(n_{ij} - n \cdot \mathsf{P}(B_iC_j))^2/n_{ij}$.

10 $I(\mathsf{P},\mathsf{P}') = \sum_{i=1}^{s}\sum_{j=1}^{t}\mathsf{P}(B_iC_j)\cdot \mathtt{ln}(\,\mathsf{P}(B_iC_j)/\mathsf{P}'(B_iC_j)))$.

References

1. Draheim, D.: Generalized Jeffrey Conditionalization - A Frequentist Semantics of Partial Conditionalization. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-319-69868-7. http://fpc.formcharts.org
2. Jeffrey, R.C.: Contributions to the theory of inductive probability. Ph.D. thesis, Princeton University (1957)
3. Jeffrey, R.C.: The Logic of Decision, 1st edn. McGraw-Hill, New York (1965)
4. Jeffrey, R.C.: The Logic of Decision, 2nd edn. University of Chicago Press, Chicago (1983)
5. Draheim, D.: Semantics of the Probabilistic Typed Lambda Calculus - Markov Chain Semantics, Termination Behavior, and Denotational Semantics. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-642-55198-7
6. Wicksell, S.D.: Some theorems in the theory of probability - with special reference to their importance in the theory of homograde correlations. Svenska Aktuarieforeningens Tidskrift, pp. 165–213 (1916)
7. Aitken, A., Gonin, H.: On fourfold sampling with and without replacement. Proc. R. Soc. Edinburgh **55**, 114–125 (1935)
8. Teicher, H.: On the multivariate poisson distribution. Skand. Aktuarietidskr. **37**, 1–9 (1954)
9. Ramsey, F.P.: The Foundations of Mathematics and other Logical Essays. Kegan, Paul, Trench, Trubner & Co., Ltd., New York (1931). Ed. by R.B. Braithwaite
10. Ramsey, F.P.: Truth and probability. In: Ramsey, F.P., Braithwaite, R. (eds.) The Foundations of Mathematics and other Logical Essays, pp. 156–198. Kegan, Paul, Trench, Trubner & Co., Ltd., New York (1931)
11. Ramsey, F.P.: Philosophical Papers. Cambridge University Press, Cambridge (1990). Ed. by D.H. Mellor
12. Jeffrey, R.C.: Subjective Probability - the Real Thing. Cambridge University Press, Cambridge (2004)
13. Jeffrey, R.C.: Probable knowledge. In: Lakatos, I. (ed.) The Problem of Inductive Logic, pp. 166–180. North-Holland, Amsterdam, New York, Oxford, Tokio (1968)
14. Levi, I.: Probability kinematics. Br. J. Philos. Sci. **18**(3), 197–209 (1967)
15. Diaconis, P., Zabell, S.: Some alternatives to Bayes's rules. Technical report No. 205, Department of Statistics, Stanford University, October 1983
16. Diaconis, P., Zabell, S.: Some alternatives to Bayes's rules. In: Grofman, B., Owen, G. (eds.) Information Pooling and Group Decision Making, pp. 25–38. JAI Press, Stamford (1986)
17. Bradley, R.: Decision Theory with a Human Face. Draft, p. 318, April 2016. http://personal.lse.ac.uk/bradleyr/pdf/DecisionTheorywithaHumanFace(indexed3).pdf (forthcoming)
18. Dietrich, F., List, C., Bradley, R.: Belief revision generalized - a joint characterization of Bayes's and Jeffrey's rules. J. Econ. Theory (forthcoming)
19. Kolmogorov, A.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Heidelberg (1933). https://doi.org/10.1007/978-3-642-49888-6
20. Kolmogorov, A.: Foundations of the Theory of Probability. Chelsea, New York (1956)
21. Kolmogorov, A.: On logical foundation of probability theory. In: Itô, K., Prokhorov, J.V. (eds.) LNM. Lecture Notes in Mathematics, vol. 1021, pp. 1–5. Springer, Heidelberg (1982). https://doi.org/10.1007/BFb0072897

22. Neyman, J.: Outline of a theory of statistical estimation based on the classical theory of probability. Philos. Trans. R. Soc. Lond. **236**(767), 333–380 (1937)
23. Neyman, J.: Frequentist probability and frequentist statistics. Synthese **36**, 97–131 (1977)
24. Weisberg, J.: Varieties of Bayesianism. In: Gabbay, D., Hartmann, S., Woods, J. (eds.) Handbook of the History of Logic, vol. 10 (2011)
25. Galavotti, M.C.: The modern epistemic interpretations of probability - logicism and subjectivism. In: Gabbay, D., Hartmann, S., Woods, J. (eds.) Handbook of the History of Logic, vol. 10, pp. 153–203. Elsevier, Amsterdam (2011)
26. Weirich, P.: The Bayesian decision-theoretic approach to statistics. In: Bandyopadhyay, P.S., Forster, M.R. (eds.) Philosophy of Statistics. Handbook of Philosophy of Science, vol. 7 (Gabbay, D.M., Thagard, P., Woods, J. general editors). North-Holland, Amsterdam, Boston Heidelberg (2011)
27. de Finetti, B.: Foresight - its logical laws, its subjective sources. In: Kyburg, H.E., Smokler, H.E. (eds.) Studies in Subjective Probability. Wiley, Hoboken (1964)
28. de Finetti, B.: Theory of Probability - A Critical Introductory Treatment. Wiley, Hoboken (2017). First issued in 1975 as a two-volume work
29. Jaynes, E.: Papers on Probability, Statistics and Statistical Physics. Kluwer Academic Publishers, Dodrecht, Boston, London (1989). Ed. by E.D. Rosenkranz
30. Jaynes, E.T.: Prior probabilities. IEEE Trans. Syst. Sci. Cybern. **4**(3), 227–41 (1968)
31. Jaynes, E.T.: Probability Theory. Cambridge University Press, Cambridge (2003)
32. Pearl, J.: Probabilistic Reasoning in Intelligent Systems - Networks of Plausible Inference, 2nd edn. Morgan Kaufmann, San Francisco (1988)
33. Pearl, J.: Causal inference in statistics - an overview. Stat. Surv. **3**, 96–146 (2009)
34. Pearl, J.: Causality - Models, Reasoning, and Inference, 2nd edn. Cambridge University Press, Cambridge (2009)
35. Deming, W.E., Stephan, F.F.: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Stat. **11**(4), 427–444 (1940)
36. Ireland, C.T., Kullback, S.: Contingency tables with given marginals. Biometrika **55**(1), 179–188 (1968)
37. Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)
38. Kullback, S., Khairat, M.: A note on minimum discrimination information. Ann. Math. Stat. **37**, 279–280 (1966)
39. Stephan, F.F.: An iterative method of adjusting sample frequency tables when expected marginal totals are known. Ann. Math. Stat. **13**(2), 166–178 (1942)
40. Neyman, J.: Contribution to the theory of the x^2 test. In: Neyman, J. (ed.) Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, pp. 239–273. University of California Press, Berkeley, Los Angeles (1946)
41. Taylor, W.F.: Distance functions and regular best asymptotically normal estimates. Ann. Math. Stat. **24**(1), 85–92 (1953)