A Frequentist Semantics for a Generalized Jeffrey Conditionalization

Dirk Draheim

Tallinn University of Technology, 5th May 2016

TTÜ

Motivation

- Partial knowledge specification

- Probability conditional on list of frequency castings

- Bayesian epistemology vs. classical, frequentist *extensio* of probability theory

$$\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_n \equiv b_n) \tag{1}$$

$$\mathsf{P}(A \mid \mathbf{B} \equiv \mathbf{b}) \tag{2}$$

Many-Valued Logics

	Lδ	Gödel logics Gk	Lukasiewicz logics Lk	Product logic Π	Post logics Pm
A ∧ B	≤ min {A,B}	min{A,B}	1-min{1,1-A-B}		
¬ A	1-A	1, if A=0 0, if A>0	1-A		1, A=0 A-(1/(m-1)), A>0
A ∨ B	≥ max {A,B}	max{A,B}	min{1,A+B}		max{A,B}
A → B	≤ min {1-A,B}	1, if A≤B B, if A>B	min{1,1-A+B}		

Jeffrey Conditionalization

Conditional Proabability

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} \tag{3}$$

Jeffrey Conditionalization – Probability Kynematics

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

Conditional Proabality as Jeffrey Conditionalization

$$\mathsf{P}(A \mid B) = \mathsf{P}(A \mid B \equiv 100\%) \tag{5}$$

$$\mathsf{P}(A \mid \overline{B}) = \mathsf{P}(A \mid B \equiv 0\%) \tag{6}$$

Frequentist Semantics of Jeffrey Conditionalization

We define:

$$\mathsf{P}^n(A \mid B \equiv b) \ =_{DEF} \ \mathsf{E}(\overline{A\ ^n} \mid \overline{B\ ^n} = b) \tag{7}$$

We have:

$$\mathsf{P}^n(A \mid B \equiv b) \ = \ \mathsf{P}(A \mid \overline{B\ ^n} = b) \tag{8}$$

Lemma 1 (Bounded F.P. Conditionalization in the Basic Jeffrey Case)
Let $b = x/y$ so that x/y is the irreducable fraction of b. For all $n = m \cdot y$ with $m \in \mathbb{N}$ we have the following:

$$\mathsf{P}^n(A \mid B \equiv b) = b \cdot \mathsf{P}(A \mid B) + (1 - b) \cdot \mathsf{P}(A \mid \overline{B}) \tag{9}$$

In particular:

$$\mathsf{P}^1(A \mid B \equiv 100\%) \ = \ \mathsf{P}(A \mid \overline{B\ ^1} = 1) \ = \ \mathsf{P}(A \mid B) \tag{10}$$

Frequentist Semantics of F.P. Conditionalization

Given $\mathbf{b} = b_1, \ldots, b_n$ so that y is the least common denominator of $\mathbf{b}$. For all $n = m \cdot y$ with $m \in \mathbb{N}$ we define bounded F.P. conditionalization:

$$\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_n \equiv b_n) \;=_{DEF}\; \mathsf{E}(\overline{A^{\;n}} \mid \overline{B_1^{\;n}} = b_1 \ldots, \overline{B_m^{\;n}}) \tag{11}$$

We have:

$$\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_n \equiv b_n) \;=_{DEF}\; \mathsf{P}(A | \overline{B_1^{\;n}} = b_1 \ldots, \overline{B_m^{\;n}}) \tag{12}$$

$$\mathsf{P}^n(A \mid \mathbf{B} \equiv \mathbf{b}) \;=_{DEF}\; \mathsf{P}(A | \overline{\mathbf{B}^{\;\mathbf{n}}} = \mathbf{b}) \tag{13}$$

We define F.P. conditionalization:

$$\mathsf{P}(A \mid \mathbf{B} \equiv \mathbf{b}) = \lim_{n' \to \infty} \mathsf{P}^n(A \mid \mathbf{B} \equiv \mathbf{b}) \quad \textit{where}\, n = n' \cdot \textit{lcd}(\mathbf{b}) \tag{14}$$

Proof of Lemma 1

$$\mathsf{P}^n(A \mid B \equiv b) \tag{15}$$

$$\mathsf{P}(A \mid \overline{B\ ^n} = b) \tag{16}$$

$$\frac{\mathsf{P}(A, \overline{B\ ^n} = b)}{\mathsf{P}(\overline{B\ ^n} = b)} \tag{17}$$

$$\frac{\mathsf{P}(AB, \overline{B\ ^n} = b)}{\mathsf{P}(\overline{B\ ^n} = b)} + \frac{\mathsf{P}(A\overline{B}, \overline{B\ ^n} = b)}{\mathsf{P}(\overline{B\ ^n} = b)} \tag{18}$$

We consider the first summand only:

$$\frac{\mathsf{P}(AB, \overline{B\ ^n} = b)}{\mathsf{P}(\overline{B\ ^n} = b)} \tag{19}$$

Proof of Lemma 1 – cont. (ii)

$$\frac{\mathsf{P}(A_1B_1, \overline{B_1 + \cdots + B_n} = b)}{\mathsf{P}(\overline{B}^{\,n} = b)} \tag{20}$$

$$\frac{\mathsf{P}(A_1B_1, \overline{B_2 + \cdots + B_n} = \frac{bn-1}{n-1})}{\mathsf{P}(\overline{B}^{\,n} = b)} \tag{21}$$

$$\frac{\mathsf{P}(A_1B_1) \cdot \mathsf{P}(\overline{B_2 + \cdots + B_n} = \frac{bn-1}{n-1})}{\mathsf{P}(\overline{B}^{\,n} = b)} \tag{22}$$

Now, due to the fact that $(B_i)_{i \in \mathbb{N}}$ is a sequence of **i.i.d** random variables, we have the following:

$$\mathsf{P}(\overline{B_2 + \cdots + B_n} = \frac{bn-1}{n-1}) = \mathsf{P}(\overline{B_1 + \cdots + B_{n-1})} = \frac{bn-1}{n-1} \tag{23}$$

Due to Eqn. (23) we can rewrite Eqn. (22), just for convenience and better readability, as follows:

$$\frac{\mathsf{P}(AB) \cdot \mathsf{P}(\overline{B}^{\,n-1} = \frac{bn-1}{n-1})}{\mathsf{P}(\overline{B}^{\,n} = b)} \tag{24}$$

Proof of Lemma 1 – cont. (iii)

Now, we have that $\mathsf{P}(AB)$ equals $\mathsf{P}(A|B) \cdot \mathsf{P}(B)$ and therefore that Eqn. (24) equals:

$$\frac{\mathsf{P}(A \mid B) \cdot \mathsf{P}(B) \cdot \mathsf{P}(\overline{B^{\,n-1}} = \frac{bn-1}{n-1})}{\mathsf{P}(\overline{B^{\,n}} = b)} \tag{25}$$

As the next crucial step, we resolve $\mathsf{P}(\overline{B^{\,n-1}} = \frac{bn-1}{n-1})$ and $\mathsf{P}(\overline{B^{\,n}} = b)$ combinatorically. We have that Eqn. (25) equals:

$$\mathsf{P}(A \mid B) \cdot \mathsf{P}(B) \cdot \frac{\binom{n-1}{bn-1} \cdot \mathsf{P}(B)^{bn-1} \cdot \mathsf{P}(\overline{B})^{n-bn}}{\binom{n}{bn} \cdot \mathsf{P}(B)^{bn} \cdot \mathsf{P}(\overline{B})^{n-bn}} \tag{26}$$

As a next step, we can cancel all occurrences of $P(B)$ and $P(\overline{B})$ from Eqn. (26) which yields the following:

$$\mathsf{P}(A \mid B) \cdot \frac{(n-1)!}{(bn-1)!(n-1-(bn-1))!} \Bigg/ \frac{n!}{(bn)!(n-bn)!} \tag{27}$$

After resolving $(n-1)!$ as $n!/n$, resolving $(bn-1)!$ to $(bn)!/(bn)$ and some further trivial transformations we have that Eqn. (27) equals:

$$\mathsf{P}(A \mid B) \cdot \frac{n!\,bn}{n\,(bn)!(n-bn)!} \cdot \frac{(bn)!(n-bn)!}{n!} \tag{28}$$

Proof of Lemma 1 – cont. (iv)

Now, after a series of further cancelations we have that Eqn. (28) equals the following:

$$b \cdot \mathsf{P}^n(A \mid B) \tag{29}$$

Similarly (omitted), it can be shown that the second summand in Eqn. (18) equals:

$$(1 - b) \cdot \mathsf{P}^n(A \mid \overline{B}) \tag{30}$$

□

Decomposition of F.P. Conditionalization

Lemma 2 (Decomposition of Bounded F.P. Conditionalization)
Given a bounded F.P. conditionalization $\mathsf{P}^n(A \mid \mathbf{B} \equiv \mathbf{b})$ *for some bound* n *and a vector of events* $\mathbf{B} = (B_i)_{\{1,\ldots,m\}}$ *for the index set* $I = \{1, \ldots, m\}$*, we have the following:*

$$\mathsf{P}^n(A \mid \mathbf{B} \equiv \mathbf{b}) = \sum_{\substack{\left(\zeta_i \in \{B_i, \overline{B_i}\}\right)_{i \in I} \\ \mathsf{P}(\bigcap_{i \in I} \zeta_i) \neq 0}} \left(\mathsf{P}(A \mid \bigcap_{i \in I} \zeta_i) \cdot \mathsf{P}^n(\bigcap_{i \in I} \zeta_i \mid \mathbf{B} \equiv \mathbf{b}) \right) \tag{31}$$

For example, in case of two conditions:

$$\begin{aligned} \mathsf{P}(A \mid B \equiv b, C \equiv c) \;=\; & \mathsf{P}(A \mid BC) \cdot \mathsf{P}(BC \mid B \equiv b, C \equiv c) \\ + \; & \mathsf{P}(A \mid B\overline{C}) \cdot \mathsf{P}(B\overline{C} \mid B \equiv b, C \equiv c) \\ + \; & \mathsf{P}(A \mid \overline{B}C) \cdot \mathsf{P}(\overline{B}C \mid B \equiv b, C \equiv c) \\ + \; & \mathsf{P}(A \mid \overline{BC}) \cdot \mathsf{P}(\overline{BC} \mid B \equiv b, C \equiv c) \end{aligned} \tag{32}$$

Computation of F.P. Conditionalization

Definition 3 (Frequency Adoption)

$$\xi_J^{l,n}(p) = \begin{cases} \frac{np-1}{n-1} & , l \in J \\ \frac{np}{n-1} & , l \notin J \end{cases} \tag{33}$$

Based on the notation for frequency adoption in Def. 3, we can define the computation of F.P. conjunctions via the following recursive equation:

$$\begin{aligned} & \mathsf{P}^n(B_1 \equiv b_1 \ldots B_m \equiv b_m) \\ = & \begin{cases} 1 & , n = 0 \\ \sum\limits_{\substack{I' \subseteq I \\ \nexists i \in I'. b_i = 0 \\ \nexists i \in \overline{I'}. b_i = 1}} \mathsf{P}\Big(\bigcap\limits_{i \in I'} B_i, \bigcap\limits_{i \in \overline{I'}} \overline{B_i}\Big) \cdot \mathsf{P}^{n-1}\Big(B_1 \equiv \xi_{I'}^{1,n}(b_1), .., B_m \equiv \xi_{I'}^{m,n}(b_m)\Big) & , n \geqslant 1 \end{cases} \end{aligned} \tag{34}$$

F.P. Conditionalization and Independency

Lemma 4 (Independence of F.P. Conditions)
Given a bounded F.P. conditionalization $\mathsf{P}^n(A \,|\, \mathbf{B} \equiv \mathbf{b})$ *for some bound* n *and a vector of mutually independent events* $\mathbf{B} = (B_i)_{\{1,\ldots,m\}}$ *for the index set* $I = \{1,\ldots,m\}$, *we have the following:*

$$\mathsf{P}^n(A \,|\, \mathbf{B} \equiv \mathbf{b}) = \sum_{I' \subseteq I} \Big(\mathsf{P}(A \,|\, \bigcap_{i \in I'} B_i, \bigcap_{i \in \overline{I'}} \overline{B_i}) \cdot \prod_{i \in I'} b_i \cdot \prod_{i \in \overline{I'}} (1 - b_i) \Big) \tag{35}$$

For example, in case of two conditions:

$$\begin{aligned} \mathsf{P}(A|B \equiv b, C \equiv c) \;=\; & \mathsf{P}(A \,|\, BC) \cdot bc \\ +\; & \mathsf{P}(A \,|\, B\overline{C}) \cdot b(1-c) \\ +\; & \mathsf{P}(A \,|\, \overline{B}C) \cdot (1-b)c \\ +\; & \mathsf{P}(A \,|\, \overline{BC}) \cdot (1-b)(1-c) \end{aligned} \tag{36}$$

Outlook – Bayesianism and Frequentism

- Jakob Bernoulli
- Bruno de Finetti
- John Maynard Keynes
- Frank P. Ramsey
- Rudolf Carnap
- Dempster-Shafer

Conclusion

- Partial knowledge specification
- Probability conditional on list of frequency castings
- Bayesian epistemology vs. classical, frequentist *extensio* of probability theory
- $\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_n \equiv b_n)$
- $\mathsf{P}(A \mid \mathbf{B} \equiv \mathbf{b})$
- In its basic case, F.P. conditionalization meets Jeffrey conditionalization
- Computation of F.P. conditionalization
- Independency and F.P. conditionalization
- F.P. conditionalization and Bayesianism vs. frequentism

Thanks a lot!

dirk.draheim@ttu.ee

Appendix

Definition 5 (Independent Random Variables) *Given to random variables* $X : \Omega \to I$ *and* $Y : \Omega \to I$*, we say that and* X *and* Y *are independent, if the following holds for all* $v \in I$ *and* $v' \in I$*:*

$$\mathsf{P}(X = v, Y = v') = \mathsf{P}(X = v) \cdot \mathsf{P}(Y = v') \tag{37}$$

Definition 6 (Identically Distributed Random Variables) *Given to random variables* $X : \Omega \to I$ *and* $Y : \Omega \to I$*, we say that and* X *and* Y *are identically distributed, if the following holds for all* $v \in I$*:*

$$\mathsf{P}(X = v) = \mathsf{P}(Y = v) \tag{38}$$

Definition 7 (Independent, Identically Distributed) *Given to random variables* $X : \Omega \to I$ *and* $Y : \Omega \to I$*, we say that and* X *and* Y *are independent identically distributed, abbreviated as* **i.i.d***, if they are both independent and identically distributed.*

Definition 8 (Sequence of i.i.d Random Variables) *Random variables* $(X_i)_{i \in \mathbb{N}}$ *are called independent identically distributed, again abbreviated as* **i.i.d***, if they are pairwise independent and furthermore identically distributed.*

Definition 9 (Matrix of i.i.d Random Variables) *Given a list* $(X_k)_{k\in R}$ *with* $R = \{1,\ldots,m\}$ *of sequences of random variables* $X_k = (X_{ki})_{i\in\mathbb{N}}$ *so that* $(X_{ki}) : \Omega \longrightarrow I$*, i.e., random variables that are organized in an* $R \times \mathbb{N}$*-matrix. These random variables are called independent identically distributed, again abbreviated as* **i.i.d***, if each row* X_k *for all* $k \in R$ *is identically distributed and furthermore, they are column-wise mutually completely independent as defined as follows. Given a designated column number* $c \in \mathbb{N}$*, numbers* $1 \leqslant n \leqslant m$*,* $1 \leqslant n' \leqslant m$*, a sequence of row indices* $i_1,\ldots,i_n$*, a sequence of row indices* $j_1,\ldots,j_{n'}$*, and a sequence of column indices* $k_1,\ldots,k_{n'}$ *so that* $k_q \neq c$ *for all* $1 \leqslant q \leqslant n'$ *we have that the following independency condition holds:*

$$\mathsf{P}(X_{i_1c},\ldots,X_{i_nc},X_{j_1k_1},\ldots,X_{j_{n'}k_{n'}}) = \mathsf{P}(X_{i_1c},\ldots,X_{i_nc}) \cdot \mathsf{P}(X_{j_1k_1},\ldots,X_{j_{n'}k_{n'}}) \tag{39}$$

A characteristic random variable is a real-valued random variable $A : \Omega \to \mathbb{R}$ that assigns only zero or one as values, i.e.:$(A = 1) \cup (A = 0) = \Omega$

A characteristic random variable stands for a Bernoulli experiment. It characterizes an event. Given an event $A \subseteq \Omega$ we define its characteristic random variable $A : \Omega \to [0,1]$ as follows:

$$A(\omega) = \begin{cases} 1 & ,\omega \in A \\ 0 & ,\omega \notin A \end{cases} \tag{40}$$

Note, that we overload the name of the event A with the name of its characteristic random variable, which does not harm, because it is always clear form the context, whether the event or the random variable is meant. We have that the value one characterizes the event A, whereas the value zero characterizes its complement $\Omega \backslash A = \overline{A}$ and, therefore, we have the following:

$$\begin{aligned} \mathsf{P}(A=1) &= \mathsf{P}(A) \\ \mathsf{P}(A=0) &= \mathsf{P}(\overline{A}) \\ \mathsf{P}(A=r) &= 0 \quad , \forall r \notin \{0,1\} \end{aligned}$$

Definition 10 (Model of the Repetition of an Event) *Given a family $(A_i)_{i \in \mathbb{N}}$ of* **i.i.d** *characteristic random variables $A_i : \Omega \longrightarrow [0,1]$. We say that $(A_i)_{i \in \mathbb{N}}$ models the repeated observation of the event $A \subset \Omega$, or just the repetition of A for short, if we have that $A = (A_1 = 1)$.*

$$(X+Y)(\omega) = X(\omega) + Y(\omega) \tag{41}$$

$$((X+Y)=r) \quad = \quad \{\,\omega \mid X(\omega)+Y(\omega) \;=\; r\} \tag{42}$$

$$\mathsf{P}((X+Y)=r) = \sum_{r_x+r_y=r} \Big(\mathsf{P}(X=r_x) + \mathsf{P}(Y=r_y) \Big) \tag{43}$$

$$X^n = \sum_{i=1}^{n} X_i \tag{44}$$

$$(r \cdot X)(\omega) = r \cdot X(\omega) \tag{45}$$

$$\overline{X^{\;n}} = 1/n \cdot X^n \tag{46}$$

$$\overline{X^{\;n}} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{47}$$

$$\begin{aligned} X^{\infty} &= \lim_{n\to\infty} X^n \\ \overline{X^{\;\infty}} &= \lim_{n\to\infty} \overline{X^{\;n}} \end{aligned} \tag{48}$$

Lemma 11 (Independency of Sums of Random Variables) *Given pairwise independent, discrete real-valued random variables* $A : \Omega \longrightarrow \mathbb{R}$, $X : \Omega \longrightarrow \mathbb{R}$, *and* $Y : \Omega \longrightarrow \mathbb{R}$, *we have that independency transports over to the sum* $X + Y$, *i.e., for all* $a \in A^{\dagger}(\Omega)$, *and* $r \in (X+Y)^{\dagger}(\Omega)$ *we have the following:*

$$\mathsf{P}(A = a, X + Y = r) = \mathsf{P}(A = a) \cdot \mathsf{P}(X + Y = r) \tag{49}$$

Lemma 12 (Independency of Multiplies of Random Variables) *Given independent, discrete real-valued random variables* $A : \Omega \longrightarrow \mathbb{R}$ *and* $X : \Omega \longrightarrow \mathbb{R}$ *as well as a real-number* $n \in \mathbb{R}$, *we have that independency transports over to the multiply* nX, *i.e., for all* $a \in A^{\dagger}(\Omega)$, *and* $r \in (nX)^{\dagger}(\Omega)$ *we have the following:*

$$\mathsf{P}(A = a, nX = r) = \mathsf{P}(A = a) \cdot \mathsf{P}(nX = r) \tag{50}$$

Corollary 13 (Independency of n-times Sums and Averages) *Given a discrete real-valued random variable* $A : \Omega \longrightarrow \mathbb{R}$, *and a list* $(X_i)_{i\in\{1,\ldots,n\}}$ *of discrete real-valued random variables* $X_i : \Omega \longrightarrow \mathbb{R}$, *so that* A *and all* X_i *are all pairwise disjoint. Then, we have that independency transports over to the* n*-times sum* X^n *as well as over to the average* $\overline{X^{\,n}}$, *i.e., for all* $a \in A^{\dagger}(\Omega)$, $r \in (X^n)^{\dagger}(\Omega)$, *and* $s \in (\overline{X^{\,n}})^{\dagger}(\Omega)$ *we have the following:*

$$\mathsf{P}(A = a, X^n = r) = \mathsf{P}(A = a) \cdot \mathsf{P}(X^n = r) \tag{51}$$

$$\mathsf{P}(A = a, \overline{X^{\,n}} = s) = \mathsf{P}(A = a) \cdot \mathsf{P}(\overline{X^{\,n}} = s) \tag{52}$$

$$A^n(\omega) = |\{i \in \{1, \ldots, n\}|\ A_i(\omega) = 1\}| \tag{53}$$

$$\mathsf{E}(X + Y \mid C) = \mathsf{E}(X|C) + \mathsf{E}(X|C) \tag{54}$$

$$\mathsf{E}(a \cdot X + b \cdot Y \mid C) = a \cdot \mathsf{E}(X|C) + b \cdot \mathsf{E}(X|C) \tag{55}$$

$$\mathsf{E}(X^n \mid C) = n \cdot \mathsf{E}(X \mid C) \tag{56}$$

$$\mathsf{E}(\overline{X^{\,n}} \mid C) = \mathsf{E}(X \mid C) \tag{57}$$

$$\mathsf{E}(\overline{X^{\,n}} \mid C) = \mathsf{P}(X \mid C) \tag{58}$$

$$\mathsf{E}(X + Y) = \mathsf{E}(X) + \mathsf{E}(X) \tag{59}$$

$$\mathsf{E}(a \cdot X + b \cdot Y) = a \cdot \mathsf{E}(X) + b \cdot \mathsf{E}(X) \tag{60}$$

$$\mathsf{E}(X^n) = n \cdot \mathsf{E}(X) \tag{61}$$

$$\mathsf{E}(\overline{X^{\,n}}) = \mathsf{E}(X) \tag{62}$$

$$\mathsf{E}(\overline{X^{\,n}}) = \mathsf{P}(X) \tag{63}$$

Theorem 14 (Weak Law of Large Numbers) *Given a countable series* $X = (X_i)_{i\in\Omega}$ *of* **i.i.d** *real-valued random variables with expectation* $\mu = E(X) = E(X_i)$*, we have that for all* $\varepsilon \in \mathbb{R}$*:*

$$\underset{n\to\infty}{lim} \mathsf{P}(\mu-\varepsilon < \overline{X\ ^n} < \mu+\varepsilon) = 1 \tag{64}$$

Theorem 15 (Strong Law of Large Numbers) *Given a countable series* $(X_i)_{i\in\Omega}$ *of* **i.i.d** *real-valued random variables with expectation* μ*, we have that for all* $\varepsilon \in \mathbb{R}$*:*

$$\mathsf{P}(\underset{n\to\infty}{lim} \overline{X\ ^n} = \mu) = 1 \tag{65}$$