A Frequentist Semantics of Partial Conditionalization

Dirk Draheim
Information Systems Group
Tallinn University of Technology

TALLINN UNIVERSITY OF TECHNOLOGY

Reykjavik, 29th Jan. 2019

Motivation

Classical Conditional Probability

$$\boxed{\mathsf{P}(A|B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}} \qquad \mathsf{P}(A|B_1 \cdots B_m) = \frac{\mathsf{P}(AB_1 \cdots B_m)}{\mathsf{P}(B_1 \cdots B_m)} \tag{1}$$

Partial Conditionalization

$$\mathsf{P}(A \mid \mathsf{P}(B_1) := b_1, \ldots, \mathsf{P}(B_m) := b_m) \ = ??? \tag{2}$$

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

Classical Conditional Probability as Special Case of Partial Conditionalization

$$\mathsf{P}(A|B_1 \cdots B_m) = \mathsf{P}(A \mid B_1 \equiv 100\%, \ldots, B_m \equiv 100\%) \tag{4}$$

$$\mathsf{P}(A|\overline{B_1} \cdots \overline{B_m}) = \mathsf{P}(A \mid B_1 \equiv 0\%, \ldots, B_m \equiv 0\%) \tag{5}$$

Jeffrey Conditionalization

Assumption / Pre-Condition: the events $B_1, \ldots, B_m$ form a partition !

$$\boxed{\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)_{\mathrm{J}} = \sum_{\substack{i=1 \\ \mathsf{P}(B_i) \neq 0}}^{m} b_i \cdot \mathsf{P}(A \mid B_i)} \tag{6}$$

$$\mathsf{P}(A \mid B \equiv b)_{\mathrm{J}} = b \cdot \mathsf{P}(A \mid B) + (1 - b) \cdot \mathsf{P}(A \mid \overline{B}) \tag{7}$$

Original Jeffrey Notation

$$PROB(A) = \sum_{\substack{i=1 \\ \mathsf{P}(B_i) \neq 0}}^{m} PROB(B_i) \cdot prob(A \mid B_i) \tag{8}$$

Richard C. Jeffrey. The Logic of Decision, 2nd edition, University of Chicago Press, 1983.

Derivation of Jeffrey's Rule in Probability Kinematics

Definition 1 (Jeffrey's Postulate) We say that *Jeffrey's postulate* holds **iff** Given an *a priori* probability P, an *a posteriori* probability $\mathsf{P}_{\mathbf{B}}$ with a list of updates $\mathbf{B} = B_1 \equiv b_1, \ldots, B_n \equiv b_n$, we have that all probabilities conditional on some event from $B_1, \ldots, B_n$ are preserved after update as long as $B_1, \ldots, B_n$ forms a partition, i.e., we have that the following holds for all events A:

$$B_1, \ldots, B_n \textit{ forms a partition} \Rightarrow \mathsf{P}_{\mathbf{B}}(A|B_i) = \mathsf{P}(A|B_i) \textit{ for all } B_i \in \mathbf{B} \tag{9}$$

Due to the law of total probability we have:

$$\mathsf{P}_{\mathbf{B}}(A) = \sum_{\substack{i=1 \\ \mathsf{P}(B_i) \neq 0}}^{m} \mathsf{P}_{\mathbf{B}}(B_i) \cdot \mathsf{P}_{\mathbf{B}}(A|B_i) \tag{10}$$

Due to Jeffrey's postulate we have:

$$\mathsf{P}_{\mathbf{B}}(A) = \sum_{\substack{i=1 \\ \mathsf{P}(B_i) \neq 0}}^{m} \overbrace{\mathsf{P}_{\mathbf{B}}(B_i)}^{b_i} \cdot \mathsf{P}(A|B_i) \tag{11}$$

Frequentist Partial (F.P.) Conditionalization

Definition 2 (Bounded F.P. Conditionalization) Given an i.i.d.sequence of multivariate characteristic random variables $(\langle A, B_1, \ldots, B_m \rangle_{(j)})_{j \in \mathbb{N}}$, a list of rational numbers $b_1, \ldots, b_m$ 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 $b_1, \ldots, b_m$. 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 $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$, as follows:

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

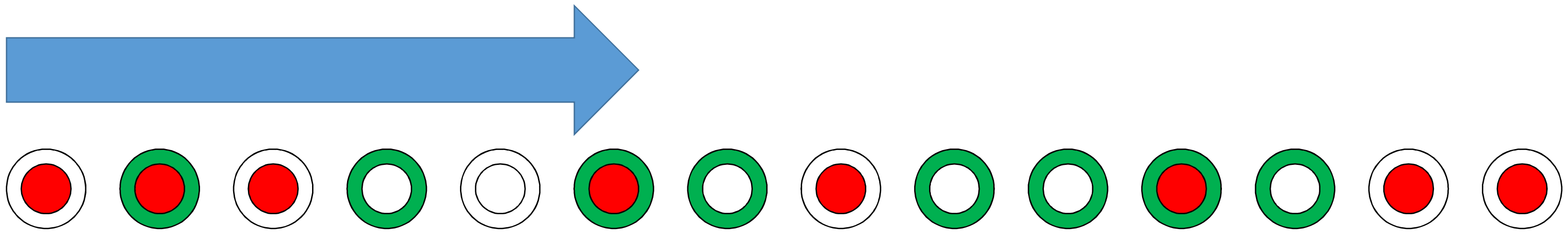
Lemma 1 (Compact Bounded F.P. Conditionalization) *Given an F.P. conditionalization* $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ *we have that the following holds:*

$$\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) = \mathsf{P}(A \mid \overline{B_1{}^n} = b_1, \ldots, \overline{B_m{}^n} = b_m) \tag{12}$$

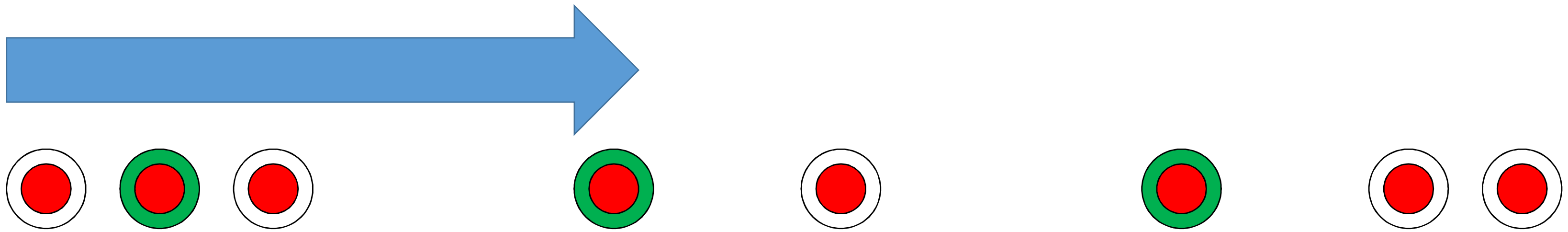
Definition 3 (F.P. Conditionalization) Given an i.i.d.sequence of multivariate characteristic random variables $(\langle A, B_1, \ldots, B_m \rangle_{(j)})_{j \in \mathbb{N}}$ and a list of rational numbers $b = b_1, \ldots, b_m$ such that $0 \leqslant b_i \leqslant 1$ for all b_i in b. We define the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

$$\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) = \lim_{k \to \infty} \mathsf{P}^{\,k \cdot lcd(b)}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) \tag{13}$$

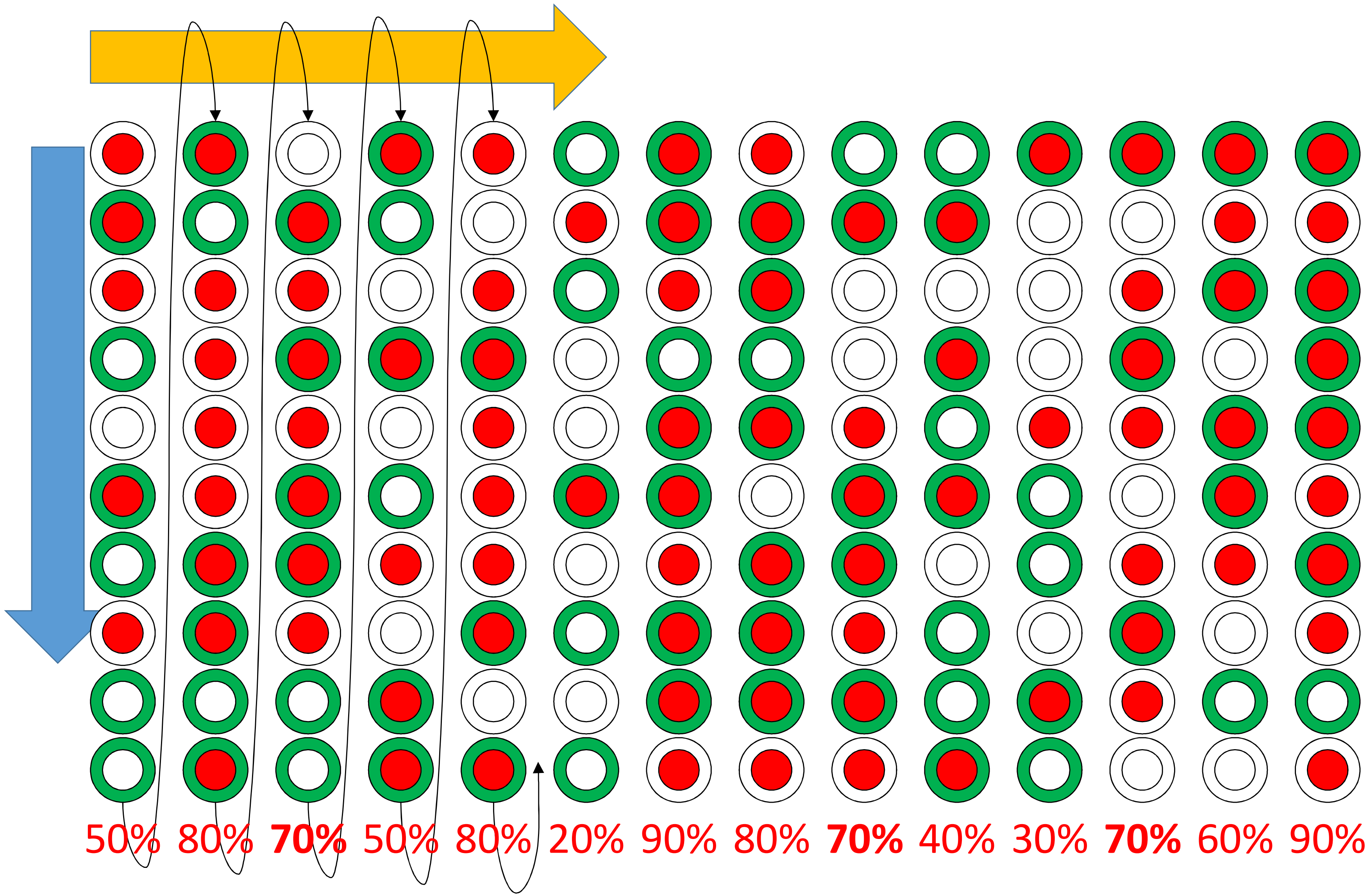
P(A|B)



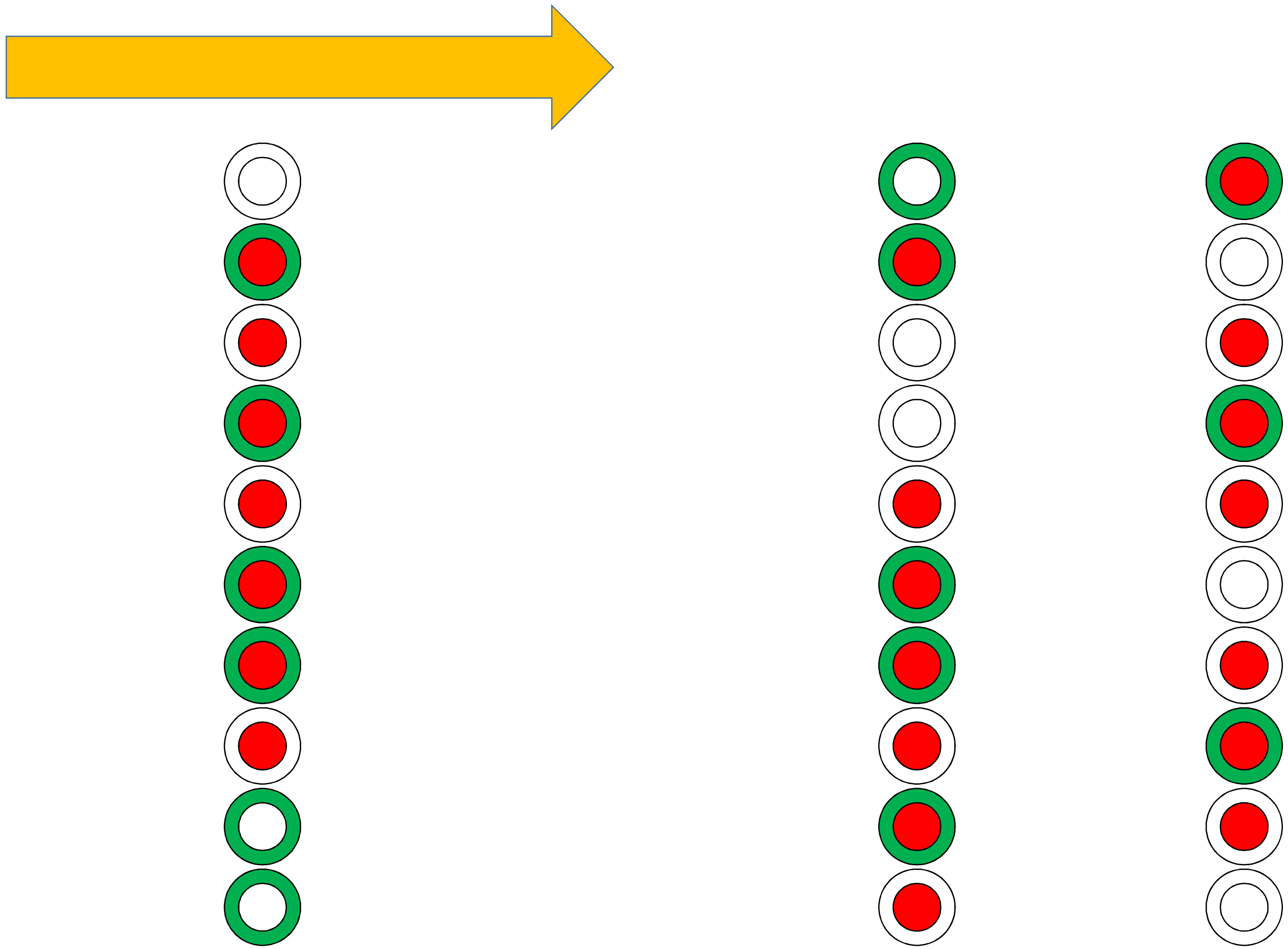
$P(A|B)$



P(A|P(B):=70%)



P(A|P(B):=70%)



50% 80% **70%** 50% 80% 20% 90% 80% **70%** 40% 30% **70%** 60% 90%

60% **50%** **30%**

F.P. Semantics of Jeffrey Conditionalization

Theorem 1 (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:*

$$\mathsf{P}(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m) = \sum_{\substack{1 \leqslant i \leqslant m \\ \mathsf{P}(B_i) \neq 0}} b_i \cdot \mathsf{P}(A \mid B_i) \tag{14}$$

Cutting Repetitions

Lemma 2 (Shortening and Adjusting) *Given a sequence of i.i.d. characteristic random variables* $(B_i)_{i\in\mathbb{N}}$*, a number of repetitions* $n \in \mathbb{N}$ *and (absolute) frequencies* $1 \leqslant k \leqslant n$ *and* $0 \leqslant k' < n$ *we have the following:*

$$(B\,,\,B^n = k) \;=\; (B\,,\,B_{(2)} + \ldots + B_{(n)} = k-1) \tag{15}$$
$$(\overline{B}\,,\,B^n = k') \;=\; (\overline{B}\,,\,B_{(2)} + \ldots + B_{(n)} = k') \tag{16}$$
$$(B\,,\,B^n = 0) \;=\; \emptyset \tag{17}$$
$$(\overline{B}\,,\,B^n = n) \;=\; \emptyset \tag{18}$$

$$\begin{aligned}
\mathsf{P}(B\,,\,B^n = k) &= \mathsf{P}(B_{(1)}\,,\,B_{(2)} + \ldots + B_{(n)} = k-1) && (19)\\
&= \mathsf{P}(B_{(1)}) \cdot \mathsf{P}(B_{(2)} + \ldots + B_{(n)} = k-1) && (20)\\
&= \mathsf{P}(B_{(1)}) \cdot \mathsf{P}(B_{(1)} + \ldots + B_{(n-1)} = k-1) && (21)\\
&= \mathsf{P}(B) \cdot \mathsf{P}(B^{n-1} = k-1) && (22)
\end{aligned}$$

$$\mathsf{P}(\overline{B}, B^n = k) = \mathsf{P}(\overline{B}) \cdot \mathsf{P}(B^{n-1} = k) \tag{23}$$

Definition 4 (Binomial Distribution) Given a Bernoulli experiment with success probability p, a number $n \in \mathbb{N}$ of experiment repetitions and a number of successes $0 \leqslant k \leqslant n$. The *binomial distribution* w.r.t. to n and p, denoted by $\mathfrak{B}_{n,p}$, determines the probability of k successes after n experiment repetitions as follows:

$$\mathfrak{B}_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{24}$$

Definition 5 (Multinomial Distribution) Given an experiment with m mutually exclusive success categories and success probabilities $p_1, \ldots, p_m$ (i.e., such that $p_1 + \cdots + p_m = 1$), a number $n \in \mathbb{N}$ of repetitions and numbers of successes $k_1, \ldots, k_m$ for each category such that $k_1 + \cdots + k_m = n$. The *multinomial distribution* w.r.t. to n and $p_1, \ldots, p_m$, denoted by $\mathfrak{M}_{n,p_1,\ldots,p_m}$ determines the probability of k_j successes in all of the success categories j after n experiment repetitions as follows:

$$\mathfrak{M}_{n,p_1,\ldots,p_m}(k_1, \ldots, k_m) = \frac{n!}{k_1! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m} \tag{25}$$

$$\begin{aligned} \mathsf{P}(A^n = k) &= \mathfrak{B}_{n,\mathsf{P}(A)}(k) \\ A_1,\ldots,A_m \textit{ is a partition} \Rightarrow \mathsf{P}(A_1^n = k_1,\ldots,A_m^n = k_m) &= \mathfrak{M}_{n,\mathsf{P}(A_1),\ldots,\mathsf{P}(A_m)}(k_1,\ldots,k_m) \end{aligned}$$

Proof of Theorem 1

Proof. We proof Eqn. (14) for all of its approximations. Due to Lemma 1 we have that $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ equals

$$\frac{\mathsf{P}(A, B_1^n = b_1 n, \ldots, B_m^n = b_m n)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)} \tag{26}$$

Due to the fact that $B_1, \ldots, B_m$ form a partition we can apply the law of total probability, to segment Eqn. (26) yielding

$$\sum_{\substack{1 \leqslant i \leqslant m \\ \mathsf{P}(B_i) \neq 0}} \frac{\mathsf{P}(A, B_i, B_1^n = b_1 n, \ldots, B_m^n = b_m n)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)} \tag{27}$$

Due to the fact that $B_1, \ldots, B_m$ forms a partition, we can rewrite Eqn. (27) as

$$\sum_{\substack{1 \leqslant i \leqslant m \\ \mathsf{P}(B_i) \neq 0}} \frac{\mathsf{P}(A, B_i, \bigcap_{\substack{0 \leqslant j \leqslant m \\ j \neq i}} \overline{B_j}, B_i^n = b_i n, \bigcap_{\substack{0 \leqslant j \leqslant m \\ j \neq i}} B_j^n = b_j n)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)} \tag{28}$$

$$\sum_{\substack{1 \leqslant i \leqslant m \\ \mathsf{P}(B_i) \neq 0}} \frac{\mathsf{P}(A, B_i, \bigcap_{\substack{0 \leqslant j \leqslant m \\ j \neq i}} \overline{B_j}, B_i^n = b_i n, \bigcap_{\substack{0 \leqslant j \leqslant m \\ j \neq i}} B_j^n = b_j n)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)} \tag{28}$$

We show that each summand in Eqn. (28) equals $b_i \cdot \mathsf{P}(A|B_i)$ for all $1 \leqslant i \leqslant m$ such that $\mathsf{P}(B_i) \neq 0$. In case $b_i = 0$ we know that $\mathsf{P}(B_i, B_i^n = b_i n) = 0$ by Eqn. (17) so that the whole summand equals zero which equals $0 \cdot \mathsf{P}(A|B_i)$ and we are done. In case $b_i \neq 0$ we can apply Eqn. (15) one time to shorten and adjust $B_i^n = b_i n$ and furthermore Eqn. (16) (m-1)-times to shorten $B_j^n = b_j n$ for all $j \neq i$ which turns the i-th summand into

$$\frac{\mathsf{P}(A, B_i, \bigcap_{\substack{0 \leqslant j \leqslant m \\ j \neq i}} \overline{B_j},\ B_{i(2)} + \cdots + B_{i(n)} = b_i n - 1, \bigcap_{\substack{0 \leqslant j \leqslant m \\ j \neq i}} B_{j(2)} + \cdots + B_{j(n)} = b_j n)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)} \tag{29}$$

$$\frac{\mathsf{P}(A, B_i, \bigcap_{\substack{0\leqslant j\leqslant m\\ j\neq i}} \overline{B_j},\ B_{i(2)} + \cdots + B_{i(n)} = b_i n - 1, \bigcap_{\substack{0\leqslant j\leqslant m\\ j\neq i}} B_{j(2)} + \cdots + B_{j(n)} = b_j n)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)} \tag{29}$$

Due to the fact that $B_1,...,B_m$ form a partition we can remove all $\overline{B_j}$ from Eqn. (29). Now, due to the fact that $(\langle A, B_1, \ldots, B_m\rangle_{(i)})_{i\in\mathbb{N}}$ is i.i.d. we can cut off $\mathsf{P}(AB_i)$ in Eqn. (29) yielding

$$\frac{\mathsf{P}(AB_i)\cdot\mathsf{P}(B_{i(2)} + \cdots + B_{i(n)} = b_i n - 1, \bigcap_{\substack{0\leqslant j\leqslant m\\ j\neq i}} B_{j(2)} + \cdots + B_{j(n)} = b_j n)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)} \tag{30}$$

According to the side condition in Eqn. (27) we can assume that $\mathsf{P}(B_i) \neq 0$. Therefore, due to $\mathsf{P}(AB_i) = \mathsf{P}(A|B_i)\cdot\mathsf{P}(B_i)$ and, furthermore, by shifting the sums we can turn Eqn. (30) into

$$\underbrace{\mathsf{P}(A|B_i)}_{\gamma_i}\cdot\underbrace{\frac{\mathsf{P}(B_i)\cdot\mathsf{P}\Big(B_i^{n-1} = b_i n - 1\,,\ \bigcap_{\substack{0\leqslant j\leqslant m\\ j\neq i}} B_j^{n-1} = b_j n\Big)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)}}_{\delta_i} \tag{31}$$

$$\underbrace{\mathsf{P}(A|B_i)}_{\gamma_i} \cdot \underbrace{\frac{\mathsf{P}(B_i) \cdot \mathsf{P}\Big(B_i^{n-1} = b_i n - 1\,,\ \bigcap_{\substack{0 \leqslant j \leqslant m \\ j \neq i}} B_j^{n-1} = b_j n\Big)}{\mathsf{P}(B_1^n = b_1 n, \ldots, B_m^n = b_m n)}}_{\delta_i} \tag{31}$$

Given Eqn. (31), it remains to be shown that $\delta_i = b_i$. Now, we can again exploit that $B_1, \ldots, B_2$ form a partition. Due to this we have that $(\langle B_1, \ldots, B_m \rangle_{(i)})_{i \in \mathbb{N}}$ determines multinomial distributions $\mathfrak{M}_{n, \mathsf{P}(B_1), \ldots, \mathsf{P}(B_m)}$ and $\mathfrak{M}_{n-1, \mathsf{P}(B_1), \ldots, \mathsf{P}(B_m)}$ in Eqn. (31). Due to this together with the Lemma's premise that $b_1 + \cdots + b_m = 1$ we can resolve factor δ_i combinatorially yielding

$$\mathsf{P}(B_i) \cdot \frac{(n-1)!}{(b_i \cdot n - 1)! \prod_{j \neq i} (b_j n)!} \cdot \mathsf{P}(B_i)^{b_i \cdot n - 1} \cdot \prod_{j \neq i} \mathsf{P}(B_j)^{b_j n} \Bigg/ \frac{n!}{\prod_{j \in I} (b_j n)!} \cdot \prod_{j \in I} \mathsf{P}(B_j)^{b_j n} \tag{32}$$

Finally, after cancellation of $\prod_{j \neq i} (b_j n)!$ and all $\mathsf{P}(B \ldots)^{\cdots}$ we arrive at

$$\frac{(n-1)!}{(b_i n - 1)!} \Big/ \frac{n!}{(b_i n)!} = \frac{n! \cdot b_i n}{n \cdot (b_i n)!} \cdot \frac{(b_i n)!}{n!} = b_i \qquad \square$$

F.P.-Jeffrey Entailment

Theorem 2 (Preservation of Conditional Probabilities w.r.t. Partitions) *Given an F.P. conditionalization* $\mathsf{P}_{\mathbf{B}}(A) = \mathsf{P}(A|B_1 \equiv b_1,..., B_m \equiv b_m)$ *such that the events* $B_1,..., B_m$ *form a partition we have that the conditional probability* $\mathsf{P}(A|B_i)$ *is preserved after update according to* $\mathbf{B}$ *for all condition events* B_i *in* $B_1, \ldots, B_m$*, i.e.:*

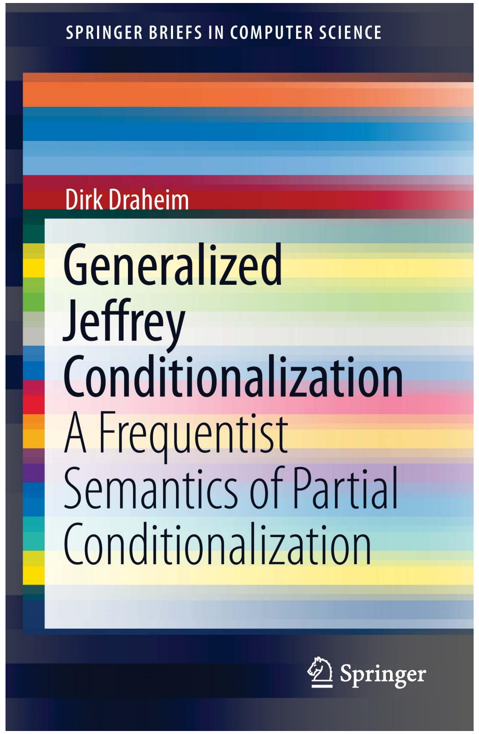
$$\mathsf{P}_{\mathbf{B}}(A|B_i) = \mathsf{P}(A|B_i) \tag{33}$$

Proof. We have that $\mathsf{P}_{\mathbf{B}}(A|B_i)$ equals $\mathsf{P}_{\mathbf{B}}(AB_i)/\mathsf{P}_{\mathbf{B}}(B_i)$. Due to the lemma's premise that $B_1, \ldots, B_m$ form a partition and Theorem 1 we have that $\mathsf{P}_{\mathbf{B}}(AB_i)/\mathsf{P}_{\mathbf{B}}(B_i)$ equals

$$\sum_{\substack{1 \leqslant j \leqslant m \\ \mathsf{P}(B_j) \neq 0}} b_j \cdot \mathsf{P}(AB_i \mid B_j) \Bigg/ \sum_{\substack{1 \leqslant j \leqslant m \\ \mathsf{P}(B_j) \neq 0}} b_j \cdot \mathsf{P}(B_i \mid B_j) \tag{34}$$

We have that Eqn. (34) equals $(b_i \cdot \mathsf{P}(AB_i|B_i))\big/(b_i \cdot \mathsf{P}(B_i|B_i))$ wich equals $\mathsf{P}(A|B_i)$. $\square$

Conclusion



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.** It analyzes the resulting partial conditionalization, **called 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 meets and generalizes Jeffrey conditionalization, i.e., from partitions to arbitrary collections of events, 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. This result is compared to Jeffrey's commutative chaining of independent updates.

The postulate of Jeffrey's probability kinematics, which is rooted in the subjectivism of Frank P. Ramsey, is found to be a consequence in our frequentist semantics. This way the book creates a link between the Kolmogorov system of probability and one of the important Bayesian frameworks. Furthermore, it shows a preservation result for conditional probabilities under the full update range and compares F.P. semantics with an operational semantics of classical conditional probability in terms of so-called conditional events. Lastly, it looks at the subjectivist notion of desirabilities and proposes a more fine-grained analysis of desirabilities a posteriori.

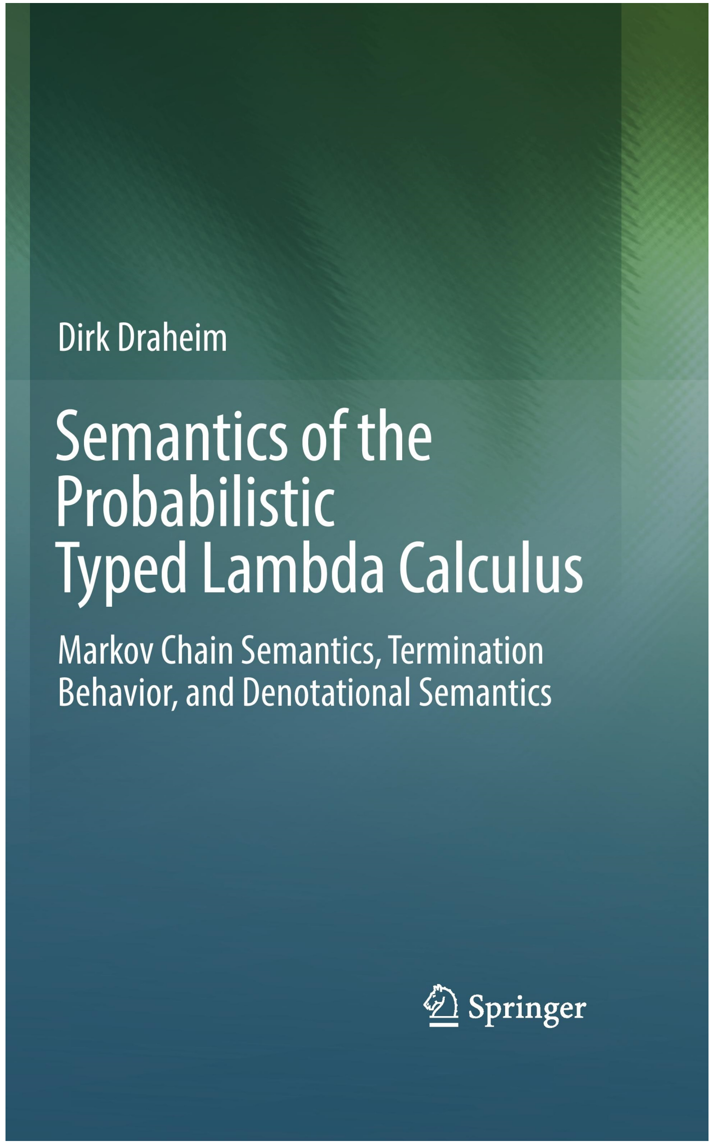
Full text PDF download from: http://fpc.formcharts.org

Thanks a lot!

Notions of Probability

Frequentism / Carnap's Probability-2	Bayesianism / Carnap's Probability-1
• *extension*: expected relative occurrence of an event in a (sufficiently) large number of repetitions of a repeatable experiment (under same conditions) • *derived (Bernoulli) extension*: relative occurence of (an individual showing) a property in a population • law of large numbers • Bernoulli's Golden Theorem • Kolomogorov • Jerzy Neyman	• *extensions*: degree of belief, preference, plausibility, validity, confirmation, uncertainty • Bruno de Finetti: Dutch book argument • Frank P. Ramsey: representation theorem • Richard C. Jeffrey: probability kinematics • Julian Jaynes: statistical reasoning, maximal entropy, agent-orientation • Judea Pearl: Bayesian networks, causality reasoning • *commong ground*: Bayes' rule, Bayesian update

Appendix



April 2017

This book takes a foundational approach to the semantics of probabilistic programming. It elaborates a rigorous Markov chain semantics for the probabilistic typed lambda calculus, which is the typed lambda calculus with recursion plus probabilistic choice. The book starts with a recapitulation of the basic mathematical tools needed throughout the book, in particular Markov chains, graph theory and domain theory, and also explores the topic of inductive definitions. It then defines the syntax and establishes the Markov chain semantics of the probabilistic lambda calculus and, furthermore, both a graph and a tree semantics. Based on that, it investigates the termination behavior of probabilistic programs. It introduces the notions of termination degree, bounded termination and path stoppability and investigates their mutual relationships. Lastly, it defines a denotational semantics of the probabilistic lambda calculus, based on continuous functions over probability distributions as domains.

Practical Exploitation

- Integration of partial conditionalization into association rule mining.

$support(A)$	$\mathsf{P}(\mathsf{A})$, $\mathsf{A}:\Omega\rightarrow\{0,1\}$
$confidence(B_1 \dots B_m \Rightarrow A)$	$\mathsf{P}(\mathsf{A} \mid \mathsf{B}_1, \dots, B_m)$
$-$	$\mathsf{E}(A\|B_1, \dots, B_m), A:\Omega\rightarrow\{i_1, \dots, i_n\} \subset \mathbb{R}$
$lift(B_1 \dots B_m \Rightarrow A) = \frac{support(A, B_1, \dots, B_m)}{support(A) \times support(B_1, \dots, B_m)}$	$\frac{\mathsf{P}(A\|B_1, \dots, B_m)}{\mathsf{P}(A)}$
$-$	$\frac{\mathsf{P}(A\|B_1, \dots, B_m)}{\mathsf{P}(A\|B_{i_1}, \dots, B_{i_k})}$
$-$	$\frac{\mathsf{E}(A\|B_1, \dots, B_m)}{\mathsf{E}(A\|B_{i_1}, \dots, B_{i_k})}$
$-$	F.P.-conditionalization-based lifts

Definition 6 (σ-Algebra) Given a set Ω, a *σ-Algebra* Σ *over* Ω is a set of subsets of Ω, i.e., $\Sigma \subseteq \mathbb{P}(\Omega)$, such that the following conditions hold true:

1) $\Omega \in \Sigma$
2) If $A \in \Sigma$ then $\Omega \backslash A \in \Sigma$
3) For all countable subsets of Σ, i.e., $A_0, A_1, A_2 \ldots \in \Sigma$ it holds true that $\bigcup_{i \in \mathbb{N}} A_i \in \Sigma$

Definition 7 (Probability Space) A *probability space* $(\Omega, \Sigma, \mathsf{P})$ consists of a set of outcomes Ω, a σ-algebra of (random) events Σ over the set of outcomes Ω and a probability function $\mathsf{P} : \Sigma \to \mathbb{R}$, also called probability measure, such that the following axioms hold true:

1) $\forall A \in \Sigma \,.\, 0 \leqslant \mathsf{P}(A) \leqslant 1$ (i.e., $\mathsf{P} : \Sigma \to [0,1]$)
2) $\mathsf{P}(\Omega) = 1$
3) (Countable Additivity): For all countable sets of pairwise disjoint events, i.e., $A_0, A_1, A_2 \ldots \in \Sigma$ with $A_i \cap A_j = \emptyset$ for all $i \neq j$, it holds true that

$$\mathsf{P}\Big(\bigcup_{i=0}^{\infty} A_i\Big) = \sum_{i=0}^{\infty} \mathsf{P}(A_i)$$

Definition 8 (Measurable Space, Measurable Function) Given two *measurable spaces* (X, Σ) and (Y, Σ'), i.e., sets X and Y equipped with a σ-algebra Σ over X and a σ-algebra Σ' over Y. A function $f : X \to Y$ is called a *measurable function*, also written as $f : (X, \Sigma) \to (Y, \Sigma')$, if for all sets $U \in \Sigma'$ we have that the inverse image $f^{-1}(U)$ is an element of Σ.

Definition 9 (Random Variable) A *random variable* X based on a probability space $(\Omega, \Sigma, \mathsf{P})$ is a measurable function $X : (\Omega, \Sigma) \to (I, \Sigma')$ with so-called *indicator set* I. The notation $(X = i)$ is used to denote the inverse image $X^{-1}(i)$ of an element $i \in I$ under f. It is usual to omit the σ-algebras in the definition of concrete random variables $X : (\Omega, \Sigma) \to (I, \Sigma')$ and specify them in terms of functions $X : \Omega \to I$ only. A random variable $X : \Omega \to I$ is called a *discrete random variable* if $X^{\dagger}(\Omega)$ is at most countable infinite.

Definition 10 (Expected Value) Given a real-valued discrete random variable $X : \Omega \to I$ with indicator set $I = \{i_0, i_1, i_2, \ldots\} \subseteq \mathbb{R}$ based on $(\Omega, \Sigma, \mathsf{P})$, the *expected value* $\mathsf{E}(X)$, or *expectation* of X (where E can also be denoted as E_{P} in so-called explicit notation) is defined as follows:

$$\mathsf{E}(X) = \sum_{n=0}^{\infty} i_n \cdot \mathsf{P}(X = i_n) \tag{35}$$

Definition 11 (Conditional Expected Value) Given a real-valued discrete random variable $X : \Omega \to I$ with indicator set $I = \{i_0, i_1, i_2, \ldots\} \subseteq \mathbb{R}$ based on a probability space $(\Omega, \Sigma, \mathsf{P})$ and an event $A \in \Sigma$, the *expected value* $\mathsf{E}(X)$ *of* X *conditional on* A (where E can also be denoted as E_P in so-called explicit notation) is defined as follows:

$$\mathsf{E}(X|A) = \sum_{n=0}^{\infty} i_n \cdot \mathsf{P}(X = i_n \mid A) \tag{36}$$

Definition 12 (Partition) Given a set M, a countable collection $B_1, B_2 \ldots$ of subsets of M is called a *partition (of M)* if $B_i \cap B_j = \emptyset$ for all $i \neq j$ and $\cup\{B_i \mid i \geqslant 1\} = M$.

Definition 13 (Independent Random Variables) Given two discrete random variables $X : \Omega \to I_1$ and $Y : \Omega \to I_2$, we say that X and Y are *independent*, if the following holds for all index values $v \in I_1$ and $v' \in I_2$:

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

Definition 14 (Pairwise Independence) Given a finite list of discrete random variables $X_1 : \Omega \to I_1$ through $X_n : \Omega \to I_n$, we say that $X_1, \ldots, X_n$ are *pairwise independent*, if X_i and X_j are independent for any two random variables $X_i \neq X_j$ from $X_1, \ldots, X_n$.

Definition 15 (Mutual Indepence) Given a finite list of discrete random variables $X_1 : \Omega \to I_1$ through $X_n : \Omega \to I_n$, we say that $X_1, \ldots, X_n$ are *(mutually) independent*, if the following holds for all index values $v_1 \in I_1$ through $v_n \in I_n$:

$$\mathsf{P}(X_1 = v_1, \ldots, X_n = v_n) = \mathsf{P}(X_1 = v_1) \times \cdots \times \mathsf{P}(X_n = v_n) \tag{38}$$

Definition 16 (Countable Independence) Given a sequence of discrete random variables $(X_i : \Omega \to I_i)_{i \in \mathbb{N}}$ we say that they are *(all) independent*, if for each finite set of indices $i_1, \ldots, i_m$ we have that $X_{i_1}, \ldots, X_{i_m}$ are mutually independent.

Definition 17 (Identically Distributed Random Variables) Given random variables $X : \Omega \to I$ and $Y : \Omega \to I$, we say that X and Y are *identically distributed*, if the following holds for all $v \in I$:

$$\mathsf{P}(X = v) = \mathsf{P}(Y = v) \tag{39}$$

Definition 18 (Independent, Identically Distributed) Given two 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 19 (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 all independent and, furthermore, all identically distributed.

Definition 20 (Multivariate Random Variable) Given a list of n so-called *marginal random variables* $X_1 : \Omega \longrightarrow I_1$ to $X_n : \Omega \longrightarrow I_n$, we define the *multivariate random variable* $\langle X_1, \ldots, X_n\rangle : \Omega \longrightarrow I_1 \times \cdots \times I_n$ for all outcomes $\omega \in \Omega$ as follows:

$$\langle X_1, \ldots, X_n\rangle(\omega) = \langle X_1(\omega), \ldots, X_n(\omega)\rangle \tag{40}$$

$$\mathsf{P}(\langle X_1, \ldots, X_n\rangle = \langle i_1, \ldots, i_n\rangle) = \mathsf{P}(X_1 = i_1, \ldots, X_n = i_n) \tag{41}$$

Corollary 1 (I.I.D. Multivariate Random Variable Marginals) *Given a sequence of i.i.d. multivariate random variables* $(\langle X_1, \ldots, X_n\rangle_i)_{i\in\mathbb{N}}$ *we have that all marginal sequences* $((X_1)_i)_{i\in\mathbb{N}}$ *through* $((X_n)_i)_{i\in\mathbb{N}}$ *are i.i.d.*

$$(X + Y)(\omega) = X(\omega) + Y(\omega) \tag{42}$$

$$((X+Y)=r) \quad = \quad \{\,\omega \mid X(\omega)+Y(\omega) \;=\; r\} \tag{43}$$

$$\mathsf{P}((X+Y)=r) = \sum_{r_x+r_y=r} \mathsf{P}(X=r_x, Y=r_y) \tag{44}$$

Convention 1 (Sum of Random Variables from First Position) Given an infinite sequence of real-valued random variables $(X_i)_{i\in\mathbb{N}}$ we use X^n to denote the sum of the first n random variables $X_1+\cdots+X_n$

Convention 2 (Sum of Random Variables from Arbitrary Position) Given an infinite sequence of real-valued random variables $(X_i)_{i\in\mathbb{N}}$ and a starting position j we use X_j^n to denote the sum $X_j+X_{j+1}+\cdots+X_{j+n-1}$. Obviously, we have that $X^n = X_1^n$.

$$\mathsf{P}(X_{i_1}+\cdots+X_{i_n}=r) = \mathsf{P}(X_{j_1}+\cdots+X_{j_n}=r) \tag{45}$$

$$\mathsf{P}(\bigcap_{1\leqslant k\leqslant m}(X_k)_{i_1}+\cdots+(X_k)_{i_n}=r_k) = \mathsf{P}(\bigcap_{1\leqslant k\leqslant m}(X_k)_{j_1}+\cdots+(X_k)_{j_n}=r_k) \tag{46}$$

$$X^0(\omega) = 0 \tag{47}$$

$$(r \cdot X)(\omega) = r \cdot X(\omega) \tag{48}$$

$$\mathsf{P}(r \cdot X = i) = \mathsf{P}(X = i/r) \tag{49}$$

$$\overline{X^n} = 1/n \cdot X^n \tag{50}$$

$$\mathsf{P}(A^n = k) = \sum_{\substack{I = \{i_1, \ldots, i_k\} \\ I' = \{i'_1, \ldots, i'_{n-k}\} \\ I \cup I' = \{1, \ldots, n\}}} \mathsf{P}(A_{(i_1)} = 1, \ldots, A_{(i_k)} = 1, A_{(i'_1)} = 0, \ldots, A_{(i'_{n-k})} = 0) \tag{51}$$

$$\mathsf{E}(X + Y \mid C) = \mathsf{E}(X|C) + \mathsf{E}(Y|C) \tag{52}$$

$$\mathsf{E}(a \cdot X + b \cdot Y \mid C) = a \cdot \mathsf{E}(Y|C) + b \cdot \mathsf{E}(X|C) \tag{53}$$

$$\forall\, 1 \leqslant i \leqslant n. \mathsf{E}(X_i|C) = \mathsf{E}(X|C) \Rightarrow \boxed{\mathsf{E}(X^n \mid C) = n \cdot \mathsf{E}(X \mid C)} \tag{54}$$

$$\forall\, 1 \leqslant i \leqslant n. \mathsf{E}(X_i|C) = \mathsf{E}(X|C) \Rightarrow \boxed{\mathsf{E}(\overline{X^n} \mid C) = \mathsf{E}(X \mid C)} \tag{55}$$

$$X \colon \Omega \to \{0, 1\}, \forall 1 \leqslant i \leqslant n. \mathsf{P}(X_i|C) = \mathsf{P}(X|C) \Rightarrow \boxed{\forall 1 \leqslant i \leqslant n. \mathsf{E}(X_i|C) = \mathsf{P}(X|C)} \tag{56}$$

$$X : \Omega \to \{0,1\}, \forall\, 1 \leqslant i \leqslant n . \mathsf{P}(X_i|C) = \mathsf{P}(X|C) \Rightarrow \boxed{\mathsf{E}(\overline{X^n} \mid C) = \mathsf{P}(X \mid C)} \tag{57}$$

Lemma 3 (Projective F.P. Conditionalizations) *Given a collection of probability specifications $B_1 \equiv b_1, \ldots, B_m \equiv b_m$ we have the following for each $1 \leqslant i \leqslant m$:*

$$\mathsf{P}(B_i \mid\ B_1 \equiv b_1, \ldots, B_m \equiv b_m) \ =\ b_i \tag{58}$$

Lemma 4 (I.I.D. Multivariate Random Variable Independencies) *Given a sequence of i.i.d. multivariate random variables $(\langle X_1, \ldots, X_n \rangle_i)_{i \in \mathbb{N}}$, a finite set $C \subset \mathbb{N}$ of column indices and a set $R_c \subseteq \{1, \ldots, n\}$ of row indices for each $c \in C$. Then, for all families of index values $((i_{\rho\kappa} : I_\rho)_{\rho \in R_\kappa})_{\kappa \in C}$ we have that the following column-wise independence holds:*

$$\mathsf{P}(\bigcap_{c \in C} \bigcap_{r \in R_c} X_{rc} = i_{rc}) = \prod_{c \in C} \mathsf{P}(\bigcap_{r \in R_c} X_{rc} = i_{rc}) \tag{59}$$

Corollary 2 (I.I.D. Multivariate Random Variable Independencies)
Given a sequence of i.i.d. multivariate random variables $(\langle X_1, \ldots, X_n \rangle_i)_{i \in \mathbb{N}}$ such that $X_1, \ldots, X_n$ are mutually independent, we have that the following holds for each index set of tuples $I \subseteq \mathbb{N} \times \mathbb{N}$:

$$\mathsf{P}(\bigcap_{\langle i,j \rangle \in I} X_{ij})) = \prod_{\langle i,j \rangle \in I} \mathsf{P}(X_{ij}) \tag{60}$$

Lemma 5 (Identical Probabilities of Target Event Repetitions) *Given an F.P. conditionalization* $\mathsf{P}^n(A \mid B_1 \equiv b_1, \ldots, B_m \equiv b_m)$ *we have that the probability of* $A_{(\sigma)}$ *conditional on the given frequency specification is equal for all repetitions* $1 \leqslant \sigma \leqslant n$*, i.e., we have for some value* ν*:*

$$\mathsf{P}(A_{(\sigma)} \mid \overline{B_1}^n = b_1, \ldots, \overline{B_m}^n = b_m) = \nu \tag{61}$$

Lemma 6 (Preservation of Independence under Aggregates) *Given* m *collections of real-valued random variables* $X_{11}, \ldots, X_{1n_1}$ *through* $X_{m1}, \ldots, X_{mn_m}$ *such that* $X_{11}, \ldots, X_{1n_1}, \ldots$ *are mutually independent, we have that the following holds true for all real numbers* $k_1, \ldots, k_m$*:*

$$\mathsf{P}(X_1^{n_1} = k_1, \ldots, X_m^{n_m} = k_m) = \mathsf{P}(X_1^{n_1} = k_1) \times \cdots \times \mathsf{P}(X_m^{n_m} = k_m) \tag{62}$$

Lemma 7 (Law of Total Probabilities) *Given a probability space* $(\Omega, \Sigma, \mathsf{P})$*, an event* $A \subseteq \Omega$ *and a countable set of events* $B_1, B_2 \ldots$ *that form a partition of* Ω*, we have that*

$$\mathsf{P}(A) = \sum_{i \geqslant 1} \mathsf{P}(AB_i) \tag{63}$$

$$\mathsf{P}(A) = \sum_{\substack{i \geqslant 1 \\ \mathsf{P}(B_i) \neq 0}} \mathsf{P}(B_i) \cdot \mathsf{P}(A|B_i) \tag{64}$$