Conditional Events, Conditional Random Variables, and the Strong Law of Large Numbers

Dirk Draheim
Information Systems Group
Tallinn University of Technology
draheim@acm.org
(with: Ahto Buldas, Cybersecurity Group, TalTech)

Tartu, 26xt April 2026

Conditional Probability and Strong Law of Large Numbers

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}$$

$$\mathsf{P}(A_B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} \quad \text{???}$$

⇒ Goodman-Nguyen-van Fraassen conditional event

$$\mathsf{P}(\lim_{n\longrightarrow\infty}\overline{A^n} = \mu_A) = 1 \qquad \mathsf{P}\Big(\big(\lim_{n\longrightarrow\infty}\overline{A^n}\big) = \mu_A\Big) = 1 \qquad A\subseteq\Omega : \mu_A = \mathsf{P}(A)$$

$$\mathsf{P}_B(\lim_{n\longrightarrow\infty}\overline{A^n} = P_B(A)) = \frac{\mathsf{P}\left(\lim_{n\longrightarrow\infty}\overline{A^n} = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}, B\right)}{\mathsf{P}(B)} = 0$$

$$\boxed{\mathsf{P}(\lim_{n\longrightarrow\infty}\overline{A^n_B} = \mathsf{P}_B(A)) = 1}$$

> *"The Kolmogorov axioms make no reference to the notion of conditional probability; indeed, KSP[*] finds this an awkward notion, really unwanted, and mentions it only reluctantly, as a seeming afterthought."*

> *"In the Kolmogorov system, conditional probability is such a foreign element that an entire book has been written (Rao, 1993)[†] trying to explain the idea of conditional probability by giving it a separate axiomatic approach!"*

Edwin T. Jaynes [‡]

[*] Kolmogorov System of Probability

[†] Malempati M. Rao. Conditional Measures and Applications, 2nd edition. Chapman and Hall, 2005 (1st edition published in 1993 by Marcel Dekker, New York)

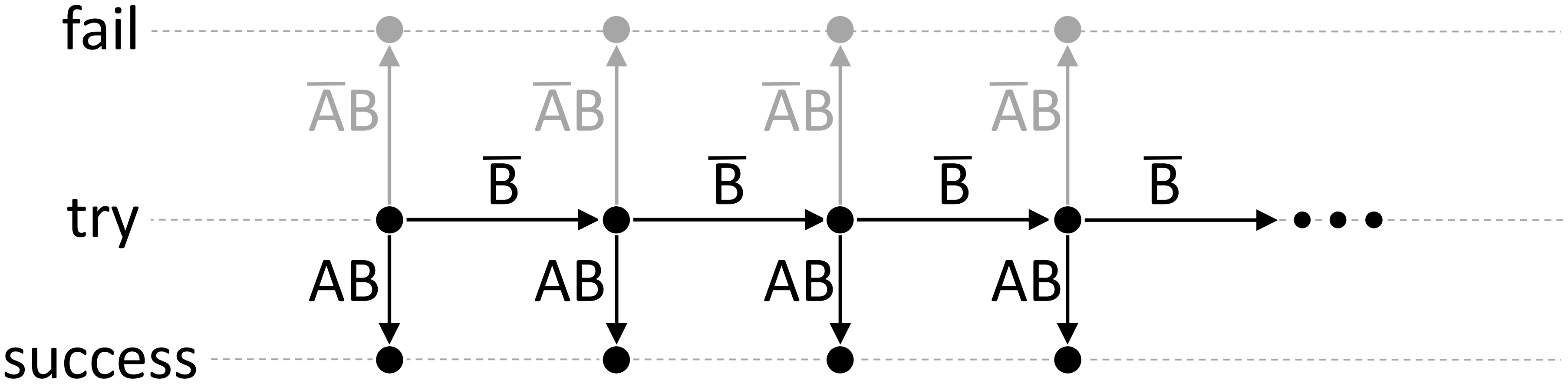
[‡] E.T. Jaynes. Probability Theory. Cambridge University Press, 2003.

Conditional Event

Informally, the conditional probability $\mathsf{P}_B(A)$ is usually explained as the *probability of event A given that event B has occurred*.

We can think of the experiment e' modeled by a conditional probability P_B as two-staged:

- An agent a_1 triggers the experiment e'.
- On behalf of this a second agent a_2 repeatedly executes the experiment e modeled by P until event B eventually occurs. Then, a_2 yields back the outcome of the last repetition of e – as result of e' – to agent a_1.



Definition 1 (Conditional Event) Given a sequence of i.i.d. multivariate characteristic random variables $(\langle A_i, B_i \rangle)_{i \in \mathbb{N}}$, we define the *conditional event* A_B as follows:

$$A_B = \bigcup_{i \in \mathbb{N}} \Big(\bigcap_{j<i} \overline{B_j},\ B_i,\ A_i \Big) \tag{1}$$

$$A_B = B_1 A_1 \cup \overline{B}_1 B_2 A_2 \cup \overline{B}_1 \overline{B}_2 B_3 A_3 \cup \overline{B}_1 \overline{B}_2 \overline{B}_3 B_4 A_4 \cup \cdots .$$

Lemma 1 (Probability of Conditional Events) *Given a sequence of i.i.d. multivariate characteristic random variables* $(\langle A_i, B_i \rangle)_{i \in \mathbb{N}}$ *so that* $\mathsf{P}(B) > 0$, *we have the following:*

$$\mathsf{P}(A_B) = \mathsf{P}_B(A) \tag{2}$$

Proof of Lemma 1

Due to the fact that all $(\cap_{j<i}\overline{B_j},\ B_i,\ A_i)$ in Eqn. (1) are disjoint and the fact that $(\langle A_i, B_i\rangle)_{i\in\mathbb{N}}$ is i.i.d. we have that $\mathsf{P}(A_B) = \mathsf{P}\Big(\bigcup_{i\in\mathbb{N}}\Big(\bigcap_{j<i}\overline{B_j},\ B_i,\ A_i\Big)\Big)$ (Def. 1) equals

$$\sum_{i\in\mathbb{N}_0} \mathsf{P}(\bigcap_{j<i}\overline{B_j})\,\mathsf{P}(A_iB_i) \tag{3}$$

Due to the fact that $(\langle A_i, B_i\rangle)_{i\in\mathbb{N}}$ and (on behalf of this also) $(B_i)_{i\in\mathbb{N}}$ are i.i.d. we have that Eqn. (3) equals

$$\sum_{i\in\mathbb{N}_0} \mathsf{P}(\overline{B})^i\,\mathsf{P}(AB) \tag{4}$$

which equals

$$\mathsf{P}(AB)\sum_{i\in\mathbb{N}_0} \mathsf{P}(\overline{B})^i \tag{5}$$

Next, a Maclaurin series expansion can be applied to the sum in Eqn. (5) so that Eqn. (5) equals $\mathsf{P}(AB)(1/(1-\mathsf{P}(\overline{B})))$ which equals $\mathsf{P}(AB)/\mathsf{P}(B)$. $\square$

Goodman-Nguyen-van Fraassen Conditional Events

$$T = (\Omega, \Sigma, \mathsf{P})$$

$$\hat{T} = (\hat{\Omega}, \hat{\Sigma}, \widehat{\mathsf{P}})$$
$$\hat{\Omega} = \Omega \times \Omega \times \Omega \times \cdots$$

$$(A|B) = \vee_{i \in \mathbb{N}_0} t_i$$
$$t_i = ((\times_{1 \leqslant i \leqslant k} \overline{A}) \times (A \wedge B) \times \Omega \times \Omega \times \cdots)$$

Notions of Probability

Carnap’s Probability-2/Frequentism	**Carnap’s Probability-1/Bayesianism**
• *standard (Kolmogorov) extension*: expected relative occurrence of an event in a (sufficiently) large number of repetitions of a repeatable experiment (under same conditions) • ***“derived” (Bernoulli) extension*: ratio of (individuals showing) a property in a population (*via “Bernoulli bridge”: assuming an auxiliary (Kolmogorov) experiment to pick individuals*)** • law of large numbers • Bernoulli’s Golden Theorem • 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

Definition 2 (n-th Conditional Event) Given a sequence of i.i.d. multivariate characteristic random variables $(\langle A_i, B_i\rangle)_{i\in\mathbb{N}}$, we define the *conditional event* $A_{B(n)}$ for each n-th first occurrence of B, also called *hitting time* n of B, where we use $\#I$ to denote the size of set I, as follows:

$$A_{B(n)} = \bigcup_{\substack{I\in\subset\mathbb{N}\\ |I|=n}} (\bigcap_{\substack{i\notin I\\ i<\mathtt{max}I}} \overline{B}_i,\ \bigcap_{iinI} B_i,\ A_{\mathtt{max}I}) \tag{6}$$

Given $n = 4$ and an index set $\{i_1, i_2, i_3, i_4\}$ such that $i_1 < i_2 < i_3 < i_4$, the event underneath the big union in Eqn. (6) has the form

$$\overline{B}_1 \cdots \overline{B}_{i_1-1}\, B_{i_1}\, \overline{B}_{i_1+1} \cdots \overline{B}_{i_2-1}\, B_{i_2}\, \overline{B}_{i_2+1} \cdots \overline{B}_{i_3-1} B_{i_3} \overline{B}_{i_3+1} \cdots \overline{B}_{i_4-1} B_{i_4}\, A_{i_4}.$$

Given $n = 1$ and an index set $\{k\}$, the event underneath the big union in Eqn. (6) has the form $B_1 \cdots B_k A_k$ and $(A_B)_1 = A_B$ equals

$$B_1A_1 \cup \overline{B}_1 B_2 A_2 \cup \overline{B}_1 \overline{B}_2 B_3 A_3 \cup \overline{B}_1 \overline{B}_2 \overline{B}_3 B_4 A_4 \cup \cdots .$$

Lemma 2 (Identical Distribution of n-th Conditional Events) *Given a sequence of i.i.d. multivariate characteristic random variables* $(\langle A_i, B_i \rangle)_{i\in\mathbb{N}}$*, we have the following for each hitting time* n*:*

$$\mathsf{P}(A_{B(n)}) = \mathsf{P}(A_{B(1)}) \tag{7}$$

Lemma 3 (Mutual Independence of Conditional Events) *Given a sequence of i.i.d. multivariate characteristic random variables* $(\langle A_i, B_i \rangle)_{i\in\mathbb{N}}$*, and mutually different indices* $n_1, \ldots, n_m$ *we have the following:*

$$\mathsf{P}(A_{B(n_1)} \cdots A_{B(n_m)}) \;=\; \mathsf{P}(A_{B(n_1)}) \times \cdots \times \mathsf{P}(A_{B(n_m)}) \tag{8}$$

Theorem 1 (Conditional Events are I.I.D.) *Given a sequence of i.i.d. multivariate characteristic random variables* $(\langle A_i, B_i \rangle)_{i\in\mathbb{N}}$*, we have that the sequence of characteristic random variables* $(A_{B(i)})_{i\in\mathbb{N}}$ *is i.i.d.*

$$\boxed{\mathsf{P}(\lim_{n\longrightarrow\infty} \overline{A_B^n} = \mathsf{P}_B(A)\,) = 1}$$

Notational Remarks

- Given a set of natural numbers S, we define the removal of its least element via a *tail function* $\tau : \mathbb{P}(\mathbb{N}) \longrightarrow \mathbb{P}(\mathbb{N})$ as $\tau(S) = S\backslash\mathtt{min}S$.
- As usual, given a charateristic random variable $X : \Omega \longrightarrow [0,1]$, we use X also to denote the event $X^{-1}(1) \subseteq \Omega$, e.g., $\omega \in B_i$ in Eqn. (9) and Eqn. (10) stands for $\omega \in B_i^{-1}(1)$, which is equivalent to $B_i(\omega) = 1$.
- As usual, given a random variable $X : \Omega \longrightarrow S$, and a Boolean mathematical expression $P[\_]$, we use $P[X]$ to denote the set $\{\omega \in \Omega \mid P[X(\omega)]\}$. For example, $(A_B)_n \in S$ in Eqn. (11) denotes the set $\{\omega \in \Omega \mid (A_B)_n(\omega) \in S\}$, which is $(A_B)_n^{-1}(S)$.
- As usual, we use (A_B) to denote $(A_B)_1$. Equally, we use A_B to denote $(A_B)_1$. We call A_B, (A_B) and $(A_B)_1$ likewise, the first repetition of A_B. We call $(A_B)_2$, $(A_B)_3$, $(A_B)_n$ etc. the second, third and n-the repetition of A_B.

Conditional Random Variables

Definition 3 (Conditional Random Variable (Basic Case – Unrepeated)) *Given a sequence of i.i.d. multivariate random variables* $(\langle A_i : \Omega \longrightarrow \mathbb{R}, B_i : \Omega \longrightarrow \{0,1\}\rangle)_{i \in \mathbb{N}}$ *, we define the* conditional random variable A_B *for an arbitrary but fixed* $\zeta \in \mathbb{R}$ *as follows:*

$$A_B(\omega) = \begin{cases} A_{\texttt{min}\{i \,|\, \omega \in B_i\}}(\omega) & , \exists i . \omega \in B_i \\ \zeta & , \textit{else} \end{cases} \tag{9}$$

Definition 4 (Conditional Random Variable (Repeated)) *Given a sequence of i.i.d. multivariate random variables* $(\langle A_i : \Omega \longrightarrow \mathbb{R}, B_i : \Omega \longrightarrow \{0,1\}\rangle)_{i \in \mathbb{N}}$ *, we define the* conditional random variable $(A_B)_n$ *for an arbitrary but fixed* $\zeta \in \mathbb{R}$ *as follows:*

$$(A_B)_n(\omega) = \begin{cases} A_{\texttt{min}\, \tau^{n-1}(\{i \,|\, B_i(\omega)=1\})}(\omega) & , \Big|\{i \,|\, \omega \in B_i\}\Big| \geqslant n \\ \zeta & , \textit{else} \end{cases} \tag{10}$$

Pre-Image of a Set under a Conditional Random Variable

Corollary 1 (Pre-Image of a Set under a Conditional Random Variable) Given a conditional random variable $(A_B)_n$ and a set $S \subseteq \mathbb{R}$ such that $0 \notin S$, we have that

$$(A_B)_n \in S = \bigcup_{\substack{I \in \mathbb{P}(\mathbb{N}) \\ |I|=n}} \left(\bigcap_{\substack{i \notin I \\ i < \mathtt{max} I}} \overline{B}_i,\ \bigcap_{i \in I} B_i,\ A_{\mathtt{max} I} \in S \right) \tag{11}$$

Proof. Immediate corollary of Def. 4. For each $\omega \in \Omega$, we have that ω belongs to the RHS event of Eqn. (10) *if and only if* it belongs to the RHS event of Eqn. (11). □

Some Examples

Given $n = 4$ and an index set $\{i_1, i_2, i_3, i_4\}$ such that $i_1 < i_2 < i_3 < i_4$, the event underneath the big union in Eqn. (11) has the form

$$\overline{B}_1 \cdots \overline{B}_{i_1-1}\, B_{i_1}\, \overline{B}_{i_1+1} \cdots \overline{B}_{i_2-1}\, B_{i_2}\, \overline{B}_{i_2+1} \cdots \overline{B}_{i_3-1} B_{i_3} \overline{B}_{i_3+1} \cdots \overline{B}_{i_4-1} B_{i_4}\, A_{i_4} \in S.$$

Given $n = 1$ and an index set $\{k\}$, the event underneath the big union in Eqn. (11) has the form $B_1 \cdots B_k A_k \in S$. If, furthermore, A is a characteristic random variable and, we have that the event underneath the big union in Eqn. (11) has the form $B_1 \cdots B_k A_k$ and $(A_B)_1$ equals the so-called conditional event A_B:

$$B_1 A_1 \cup \overline{B}_1 B_2 A_2 \cup \overline{B}_1 \overline{B}_2 B_3 A_3 \cup \overline{B}_1 \overline{B}_2 \overline{B}_3 B_4 A_4 \cup \cdots .$$

Conditional Expected Values

Discrete Case, $A : \Omega \longrightarrow \{x_1, \ldots, x_n\} \subset \mathbb{R}$:

$$\mathsf{E}(A) = \sum_{i=n}^{n} x_i \cdot \mathsf{P}(A = x_i) \tag{12}$$

Continuous Case, $A : \Omega \longrightarrow \mathbb{R}_0^+$:

$$\mathsf{E}(A) \overset{\mathtt{DEF}}{=} \int_0^\infty x f(x)\, dx = \int_0^\infty x \frac{\mathsf{P}(A < x)}{\delta x}\, dx \tag{13}$$

$$\mathsf{E}(A) \overset{\mathtt{DEF}}{=} \int\limits_\Omega A \, \mathsf{dP} \overset{\mathtt{DEF}}{=} \int_0^\infty \mathsf{P}(A > y)\, dy \tag{14}$$

$$\mathsf{E}_B(A) \overset{\mathtt{DEF}}{=} \frac{1}{\mathsf{P}(B)} \int\limits_B A \, \mathsf{dP} \tag{15}$$

$$\int\limits_B A \, \mathsf{dP} \overset{\mathtt{DEF}}{=} \int_0^\infty \mathsf{P}(B, A > y)\, dy \tag{16}$$

Theorem 2 (Conditional Expected Value) *Given a conditional random variable A_B, we have that*

$$\mathsf{E}(A_B) = \mathsf{E}_B(A) \qquad , \textit{if } \mathsf{P}(B) > 0 \tag{17}$$

$$\mathsf{E}(A_B) = 0 \qquad , \textit{if } \mathsf{P}(B) = 0 \tag{18}$$

Proof of Theorem 2

Without loss of generality, we choose $\zeta = 0$. Without loss of generality, we proof only the case that $A : \Omega \longrightarrow \mathbb{R}_0^+$. Due to Eqn. (15), we have that $\mathsf{E}(A_B)$ equals

$$\int\limits_{\Omega} A_B \,\mathsf{dP} \tag{19}$$

Due to Eqn. (16), we have that Eqn. (19) equals the (improper) Riemann integral

$$\int_0^\infty \mathsf{P}(A_B > y)\, dy \tag{20}$$

Due to Corollary 1, we have that Eqn. (20) equals

$$\int_0^\infty \mathsf{P}\Bigg(\bigcup_{n\in\mathbb{N}} \Big(\underset{j<n}{\cap} \overline{B}_j, B_n,\ A_n > y \Big) \Bigg)\, dy \tag{21}$$

All events underneath the big union in Eqn. (21) are disjoint. Therefore, and due to the countable additivity of P, we have that Eqn. (21) equals

$$\int_0^\infty \sum_{n\in\mathbb{N}} \mathsf{P}\Big(\underset{j<n}{\cap} \overline{B}_j, B_n,\ A_n > y \Big)\, dy \tag{22}$$

In case that $\mathsf{P}(B) = 0$, we can see that Eqn. (22) equals zero, which proofs Eqn. (18). Now, due to the fact that $\langle A_i, B_i \rangle$ is i.i.d., we have that (22) equals

$$\int_0^\infty \sum_{n \in \mathbb{N}} \Big(\mathsf{P}\Big(\bigcap_{j<n} \overline{B}_j \Big) \, \mathsf{P}\Big(B_n, \; A_n > y \Big) \Big) \, dy \tag{23}$$

Next, due to the fact that $\langle A_i, B_i \rangle_{i \in \mathbb{N}}$ is i.i.d., we have that $(B_i)_{i \in \mathbb{N}}$ is i.i.d. Now, due to the fact that both $\langle A_i, B_i \rangle_{i \in \mathbb{N}}$ and $(B_i)_{i \in \mathbb{N}}$ are i.i.d., we have that (23) equals

$$\int_0^\infty \sum_{n \in \mathbb{N}} \Big(\mathsf{P}(\overline{B})^n \, \mathsf{P}\Big(B, A > y \Big) \Big) \, dy \tag{24}$$

Now, we have that Eqn. (24) equals

$$\sum_{n \in \mathbb{N}} \mathsf{P}(\overline{B})^n \int_0^\infty \mathsf{P}\Big(B, A > y \Big) \, dy \tag{25}$$

Due to Maclaurin series expansion, we have that the series in Eqn. (25) equals $1/(1 - \mathsf{P}(\overline{B}))$, which equals $1/\mathsf{P}(B)$. Finally, due to Eqn. (16), we can rewrite the Riemann integral in Eqn. (25) as a Lebesgue integral, so that Eqn. (25) equals

$$\frac{1}{\mathsf{P}(B)} \int_B A \, \mathsf{dP} = \mathsf{E}_B(A) \tag{26}$$

□

Conditional Random Variables are i.i.d.

Lemma 4 (Conditional Random Variables are Identically Distributed) *Given a conditional random variable* A_B*, we have that* $((A_B)_n)_{n\in\mathbb{N}}$ *is identically distributed, in particular, we have the following for each* $n \in \mathbb{N}$ *and each real number* $x \in \mathbb{R}$*:*

$$\mathsf{P}\Big((A_B)_n < x\Big) = \mathsf{P}_B(A < x) \quad ,\textit{if } \mathsf{P}(B){>}0 \tag{27}$$

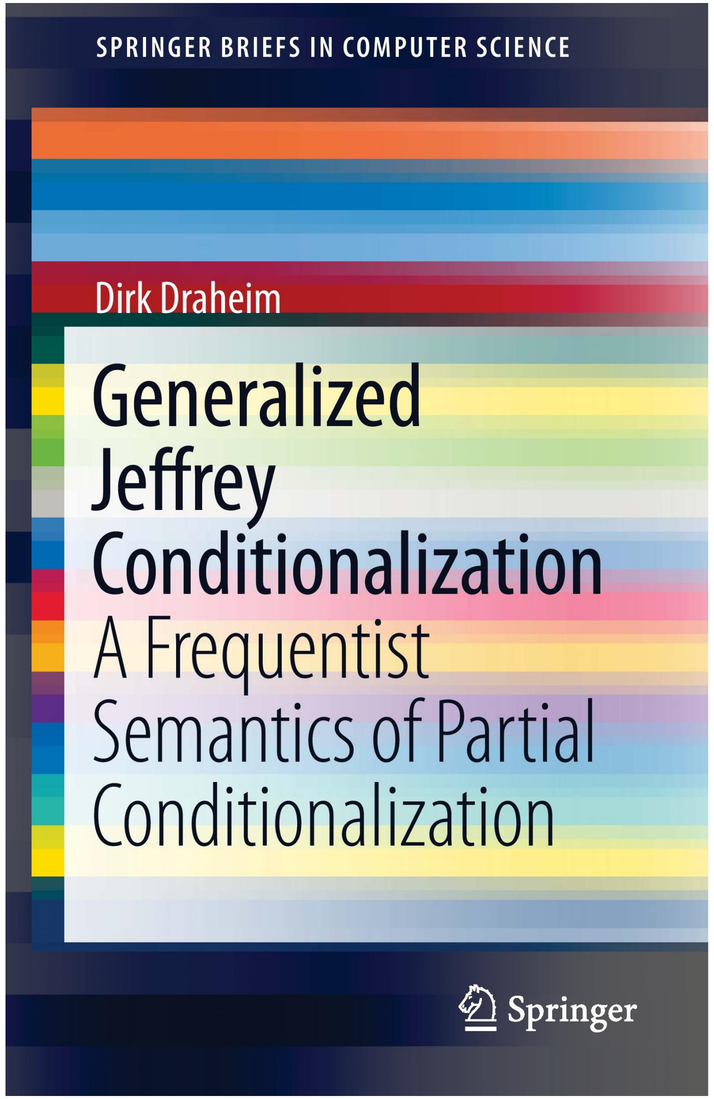
$$\mathsf{P}((A_B)_n < x\Big) = 0 \quad ,\textit{if } \mathsf{P}(B){=}0 \tag{28}$$

Lemma 5 (Conditional Random Variables are Totally Independent) *Given a conditional random variable* A_B*, we have that for each ordered set of* ρ *repetition indices* ß$_1 < \ldots <$ ß$_\rho$ *combined with any sequence of* ρ *real numbers* $(x_r)_{1\leqslant r\leqslant\rho}$*, we have that*

$$\mathsf{P}\Big(\bigcap_{1\leqslant r\leqslant\rho}\Big((A_B)_{i_r} \leqslant x_r\Big)\Big) = \prod_{1\leqslant r\leqslant\rho} \mathsf{P}_B((A_B)_{i_r} \leqslant x_r\Big) \tag{29}$$

Strong Law of Large Numbers for Conditional Probabilities

$$\boxed{\mathsf{P}\big(\lim_{n\longrightarrow\infty} \overline{A_B^n} = \mathsf{E}_B(A)\,\big) = 1}$$



SPRINGER BRIEFS IN COMPUTER SCIENCE

Dirk Draheim

Generalized Jeffrey Conditionalization

A Frequentist Semantics of Partial Conditionalization

Springer

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

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.

Conclusion

$$A_{B(n)} = \bigcup_{\substack{I \in \subset \mathbb{N} \\ |I|=n}} \Big(\bigcap_{\substack{i \notin I \\ i < \mathtt{max} I}} \overline{B}_i,\ \bigcap_{iinI} B_i,\ A_{\mathtt{max} I} \Big) \qquad (A_B)_n(\omega) = A_{\mathtt{min}\,\tau^{n-1}(\{i \,|\, B_i(\omega)=1\})}(\omega)$$

$$\boxed{\mathsf{P}(\,\mathit{lim}_{n \longrightarrow \infty} \overline{A_B^n} = \mathsf{E}_B(A)\,) = 1}$$

Preprint: https://www.researchgate.net/publication/322117507

Thanks a lot for your attention!

draheim@acm.org