

Abstract. Probabilistic inference forms lead from point probabilities of the premises to interval probabilities of the conclusion. The probabilistic version of Modus Ponens, for example, licenses the inference from $P(A) = \alpha$ and $P(B|A) = \beta$ to $P(B) \in [\alpha\beta, \alpha\beta + 1 - \alpha]$. We study *generalized* inference forms with three or more premises. The generalized Modus Ponens, for example, leads from $P(A_1) = \alpha_1, \dots, P(A_n) = \alpha_n$ and $P(B|A_1 \wedge \dots \wedge A_n) = \beta$ to an according interval for $P(B)$. We present the probability intervals for the conclusions of the generalized versions of Cut, Cautious Monotonicity, Modus Tollens, Bayes' Theorem, and some SYSTEM O rules. Recently, Gilio has shown that generalized inference forms “degrade”—more premises lead to less precise conclusions, i.e., to wider probability intervals of the conclusion. We also study Adam's probability preservation properties in generalized inference forms. Special attention is devoted to zero probabilities of the conditioning events. These zero probabilities often lead to different intervals in the coherence and the Kolmogorov approach.

Keywords: Probability logic, Generalized inference forms, Degradation, Probability preservation, Coherence.

1. Introduction

While logic studies the propagation of truth values from premises to conclusions, probability logic studies the propagation of probabilities from premises to conclusions. In probability logic Modus Ponens, for example, has the form shown on the left hand side of Table 1. On the right hand side the *generalized* probabilistic Modus Ponens is shown. In probability logic a conditional $A \Rightarrow B$ is represented by the “conditional event” $B|A$. The probabilities of the premises are point probabilities. This assessment is assumed to be coherent. Usually, the inferred probability of the conclusion is an interval probability.

Special Issue Logic and Probability: Reasoning in Uncertain Environments
Edited by Matthias Unterhuber and Gerhard Schurz

Received November 2, 2012

Table 1. Probabilistic Modus Ponens with one categorical premise (left) and generalized probabilistic Modus Ponens with n categorical premises (right)

$P(E_1) = \alpha_1$	
\dots	
$P(E) = \alpha$	$P(E_n) = \alpha_n$
$\frac{P(H E) = \beta}{P(H) \in [\delta', \delta'']}$	$\frac{P(H E_1 \wedge \dots \wedge E_n) = \beta_n}{P(H) \in [\delta'_n, \delta''_n]}$

Below we study, for different generalized inference forms, the behavior of the interval $[\delta'_n, \delta''_n]$ for an increasing number n of premises. We review results recently obtained for generalized probabilistic inference forms [6, 10, 14, 15]. In these inference forms a *degradation* is observed. The width of the probability interval of the conclusion increases as the number n of premises increases. Thus, more premises lead to less precise conclusions. Figure 1 shows a numerical example for the degradation of Modus Ponens. In most inference forms even an “ultimate” degradation occurs: Already after the addition of a small number of premises, the interval of the conclusion becomes the non-informative interval $[0, 1]$. This is a consequence of the fact, that already for small n the lower bound of the conjunction $P(\bigwedge_{i=1}^n E_i)$ may be zero.

Because the lower bound of the conjunction probability—even for a relative small number of conjuncts—becomes zero, the conditioning events may have zero probabilities. However, in the Kolmogorov approach conditional probability is undefined in this case. The Kolmogorov approach is therefore not appropriate to investigate generalized inference forms. The case where

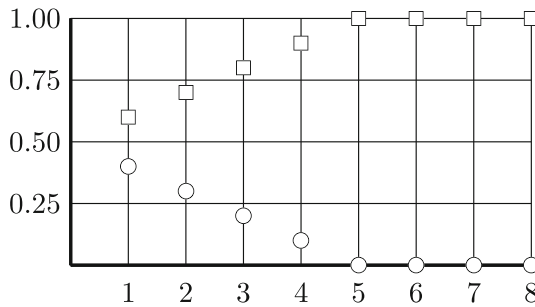


Figure 1. Degradation of Modus Ponens: lower (○) and upper (□) bounds of $P(H)$ (on the y -axes) for increasing numbers of categorical premises n (on the x -axes); premise probabilities $P(E_i) = 0.8$, for $i = 1, \dots, 8$, and $P(H | \bigwedge_{i=1}^n E_i) = 0.5$

the conditioning event has zero probability can, however, be treated in the coherence approach of de Finetti [4]. As a consequence, the Kolmogorov and the coherence approach lead to different interval probabilities for the conclusion of generalized inference forms.

As already mentioned, in many generalized inference forms the interval for the conclusion is getting wider as the number of premises increases and the interval $[0, 1]$ is obtained after a certain number of premises is added. In probabilistically valid inference forms, however, the probability of the premises of an inference form is preserved to its conclusion. Are generalized inference forms consequently probabilistically invalid? Different inference forms preserve the probability of its premises to its conclusion to different degrees. Adams distinguished four *preservation properties* [1]. Each of these preservation properties determines a consequence operation. An inference form is valid with respect to such a consequence operation, if and only if, it satisfies the corresponding preservation property. We can establish whether an inference form satisfies a preservation property by considering the lower probability of its conclusion. Well-known examples are SYSTEM P [11], which is associated with probability one-preservation and SYSTEM O [8, 9], which is closely connected with minimum probability preservation. Modus Ponens, for instance, is probability one preserving and consequently SYSTEM P valid. This can immediately be seen by considering the lower bound of the interval of the conclusion of Modus Ponens. If $P(A) = \alpha = 1$ and $P(B|A) = \beta = 1$, then $P(B) \geq \alpha\beta = 1$. It is important to note that, since they yield different intervals for the conclusion of inference forms, the Kolmogorov and the coherence approach validate different inference forms.

2. Coherent Conditional Probability

For the treatment of conditioning events with zero probabilities in generalized inference forms, we employ the coherence approach of probability theory [3, 4]. While in the Kolmogorov approach conditional probability is defined by a ratio of two (unconditional) probabilities, it is a primitive concept in the coherence approach.

Let L be a Boolean algebra (i.e., L is closed under \neg, \wedge, \vee), let \models denote the classical consequence operation and let \top be the sure event.

DEFINITION 1. A mapping $P : L \rightarrow [0, 1]$ is a (finitely additive) *one-place probability function* iff for all $A, B \in L$

1. $P(\top) = 1$,
2. $P(A \vee B) = P(A) + P(B)$, if $\models \neg(A \wedge B)$.

DEFINITION 2. Let $P' : L \rightarrow [0, 1]$ be a one-place probability function and $T_0 = \{B \in L : P'(B) \neq 0\}$. A mapping $P : L \times T_0 \rightarrow [0, 1]$ is the (finitely additive) *Kolmogorov conditional probability* associated with P' iff for all $A \in L$ and $B \in T_0$

$$P(A|B) = \frac{P'(A \wedge B)}{P'(B)}.$$

For every Kolmogorov conditional probability, if $P(B) = P(B|\top) = P'(B) = 0$, then $P(A|B)$ is undefined. The coherence approach does not exclude the case of zero probability of the conditioning event.

Let L be a Boolean algebra, $T \subseteq L$ such that T is closed under disjunction and $T^0 = T \setminus \{A \in T : A \text{ inconsistent}\}$ be a set of conditioning events.

DEFINITION 3. A mapping $P : L \times T^0 \rightarrow [0, 1]$ is a (finitely additive) *conditional probability* on $L \times T^0$ iff

1. $P(H|H) = 1$, for every $H \in T^0$,
2. $P(\cdot|H)$ is a (one-place) probability function on L for every $H \in T^0$,
3. $P(E \wedge A|H) = P(E|H)P(A|(E \wedge H))$, for $A, E \in L, H, E \wedge H \in T^0$. (see, e.g., [3])

Let L', T' be arbitrary sets of events.

DEFINITION 4. A mapping $P : L' \times T' \rightarrow [0, 1]$ is *coherent* iff there exist a Boolean algebra $L \supseteq L'$ and a set $T \subseteq L$ closed under disjunction with $T' \subseteq T^0$ such that P can be extended to a conditional probability on $L \times T^0$.

There are several advantages of the coherence approach. First, in the Kolmogorov approach, $P(A|B)$ is defined by $\frac{P(A \wedge B)}{P(B)}$. Knowledge of the marginal probabilities $P(A \wedge B)$ and $P(B)$, however, is not required to assess conditional probabilities in the coherence approach [3]. Second, while in the Kolmogorov approach $P(A|B)$ is undefined if $P(B) = 0$, in the coherence approach conditioning on (consistent) events with zero probability is possible. $P(\cdot|B)$ is a one-place probability function even if $P(B) = 0$. As a consequence, for instance, probability one can be updated in the light of events with zero probability, i.e., it is not necessarily the case that $P(A|B) = 1$, if $P(A) = 1$ [3].

The interval of the coherent probability values for the conclusion of an inference form can be determined by solving sequences of linear systems. This is the content of Theorem 5 below which provides an alternative characterization of coherence [3, p. 81]. Let $P(E_1|H_1), \dots, P(E_n|H_n)$ be a probability assessment. If $H_i = \top$, then we write $P(E_i)$ instead of $P(E_i|H_i)$.

Table 2. Constituents C_1, \dots, C_8 and their probabilities x_1, \dots, x_8 for $n = 3$ events

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	Probability
E	1	1	1	1	0	0	0	0	$x_1 + x_2 + x_3 + x_4$
F	1	1	0	0	1	1	0	0	$x_1 + x_2 + x_5 + x_6$
G	1	0	1	0	1	0	1	0	$x_1 + x_3 + x_5 + x_7$
$P(C_i)$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	

A possible outcome or a *constituent* is a logically consistent conjunction of the form $\pm E_1 \wedge \dots \wedge \pm E_n \wedge \pm H_1 \wedge \dots \wedge \pm H_n$, where $\pm A \in \{A, \neg A\}$ for all events A . If the $2n$ events are logically independent, then there are 2^{2n} constituents $C_1, \dots, C_{2^{2n}}$. The probability of an event E is the sum of the probabilities of the constituents C_r verifying it, i.e., $C_r \models E$. Table 2 shows our notation in the case of three events E, F, G .

THEOREM 5. (Coletti and Scozzafava [3, p. 81, Theorem 4]) *An assessment $P(E_1|H_1), \dots, P(E_n|H_n)$ is coherent iff there exists a sequence of compatible systems, with unknowns $x_r^\alpha \geq 0$,*

$$\mathcal{S}_\alpha = \begin{cases} \sum_{C_r \models E_i \wedge H_i} x_r^\alpha = P(E_i|H_i) \sum_{C_r \models H_i} x_r^\alpha \\ \left[\text{if } \sum_{C_r \models H_i} x_r^{\alpha-1} = 0, \alpha \geq 1 \right] \\ \sum_{C_r \models H_0^\alpha} x_r^\alpha = 1 \end{cases} \quad (i = 1, \dots, n)$$

with $\alpha = 0, 1, \dots, k \leq n$, where $H_0^0 = H_1 \vee \dots \vee H_n$ and H_0^α denotes, for $\alpha \geq 1$, the union of the H_i such that $\sum_{C_r \models H_i} x_r^{\alpha-1} = 0$.

Let P be a coherent extension of the assessment $P(E_1|H_1), \dots, P(E_n|H_n)$. Then any given solution (x_r^α) of the System \mathcal{S}_α can be interpreted as a coherent extension of the initial assessment to the family $\{C_r|H_0^\alpha : C_r \models H_0^\alpha\}$ [2].

To improve readability, we write x_i instead of x_i^0 and y_i instead of x_i^1 .

EXAMPLE 1. Consider for example Predictive Inference. The premises are $P(E) = \gamma_1, P(F) = \gamma_2, P(G) = \gamma_3$.

If $P(E \wedge F) = x_1 + x_2 > 0$, then we obtain the lower (upper) bound for the predictive probability $P(G|E \wedge F)$ by minimizing (maximizing) the objective function $\frac{x_1}{x_1+x_2}$ in the system \mathcal{S}_0

$$\begin{aligned}
 x_1 + x_2 + x_3 + x_4 &= \gamma_1 \\
 x_1 + x_2 + x_5 + x_6 &= \gamma_2 \\
 x_1 + x_3 + x_5 + x_7 &= \gamma_3 \\
 P(G|E \wedge F)(x_1 + x_2) &= x_1 \\
 \sum_{i=1}^8 x_i &= 1, \quad x_i \geq 0.
 \end{aligned}$$

Solving the linear system shows that

$$P(G|E \wedge F) \in \left[\max \left\{ 0, \frac{\gamma_1 + \gamma_2 + \gamma_3 - 2}{\gamma_1 + \gamma_2 - 1} \right\}, \min \left\{ 1, \frac{\gamma_3}{\gamma_1 + \gamma_2 - 1} \right\} \right].$$

If $P(E \wedge F) = x_1 + x_2 = 0$, then H_0^1 in Theorem 5 is $E \wedge F$. The system \mathcal{S}_1 is consequently given by

$$\begin{aligned}
 P(G|E \wedge F)(y_1 + y_2) &= y_1 \\
 y_1 + y_2 &= 1, \quad y_1, y_2 \geq 0.
 \end{aligned}$$

Solving \mathcal{S}_1 shows that, if $P(E \wedge F) = 0$, then $P(G|E \wedge F)$ can attain any value in $[0, 1]$. Note that in the Kolmogorov approach no corresponding result is obtained as in this case $P(G|E \wedge F)$ is undefined.

3. Probability Intervals for Generalized Inference Forms

In this section, we collect results for probabilistic versions of important generalized inference forms [6, 14]. We analyze these inference forms with respect to degradation. If some of the conditioning events have zero probability, we often obtain different intervals for the coherence and the Kolmogorov approach. In the coherence approach a proper treatment of this case is possible, so that the probability of the conclusion is always a closed interval. In the Kolmogorov approach, we obtain in many cases half-open, open, or no intervals at all.

For the remainder of the paper, we suppose that P is a coherent conditional probability.

3.1. And Rule

THEOREM 6. *If $P(E_i|H) = \alpha_i$, for $i = 1, \dots, n$, then*

$$P \left(\bigwedge_{i=1}^n E_i | H \right) \in \left[\max \left\{ 0, \sum_{i=1}^n \alpha_i - (n - 1) \right\}, \min \{ \alpha_i \} \right].$$

The lower bound of $P(\bigwedge_{i=1}^{n+1} E_i|H)$ is less than or equal to that of $P(\bigwedge_{i=1}^n E_i|H)$. Equality holds for lower bounds greater than zero if and only if $P(E_{n+1}|H) = \alpha_{n+1} = 1$. Moreover, if $n \geq \sum_{i=1}^n \alpha_i + 1$, then the lower bound of $P(\bigwedge_{i=1}^n E_i|H)$ is 0. We shall soon see that these properties of the conjunction cause the degradation of many other inferences.

3.2. Cautious Monotonicity

The generalized version of SYSTEM P rule Cautious Monotonicity is given by

THEOREM 7. (Gilio [6]) *If $P(E_i|E_0) = \alpha_i$, for $i = 1, \dots, n + 1$, then $P(E_{n+1}|E_0 \wedge E_1 \wedge \dots \wedge E_n) \in [\gamma', \gamma'']$, with*

$$\gamma' = \begin{cases} \max \left\{ 0, \frac{\sum_{i=1}^{n+1} \alpha_i - n}{\sum_{i=1}^n \alpha_i - (n-1)} \right\} & \text{if } \sum_{i=1}^n \alpha_i - (n-1) > 0 \\ 0 & \text{if } \sum_{i=1}^n \alpha_i - (n-1) \leq 0 \end{cases}$$

$$\gamma'' = \begin{cases} \min \left\{ 1, \frac{\alpha_{n+1}}{\sum_{i=1}^n \alpha_i - (n-1)} \right\} & \text{if } \sum_{i=1}^n \alpha_i - (n-1) > 0 \\ 1 & \text{if } \sum_{i=1}^n \alpha_i - (n-1) \leq 0 \end{cases}$$

REMARK 8. Suppose that $\alpha_{n+1} < 1$. If P is a Kolmogorov probability and $\sum_{i=1}^n \alpha_i - (n-1) \leq 0$, then the upper bound 1 cannot be attained. $P(E_{n+1}|E_0 \wedge E_1 \wedge \dots \wedge E_n) = 1$, if and only if, $P(E_0 \wedge E_1 \wedge \dots \wedge E_n) = P(E_0 \wedge E_1 \wedge \dots \wedge E_n \wedge E_{n+1})$. This requires that $P(E_0 \wedge E_1 \wedge \dots \wedge E_n) = 0$ and hence that $P(E_{n+1}|E_0 \wedge E_1 \wedge \dots \wedge E_n)$ is undefined.

Cautious Monotonicity degrades. As the number of premises increases, the width of the interval of the conclusion increases. Furthermore, if $n \geq \sum_{i=1}^n \alpha_i + 1$, then $P(E_{n+1}|E_0 \wedge E_1 \wedge \dots \wedge E_n) \in [0, 1]$.

3.3. Cut

The generalized version of SYSTEM P rule Cut is given by the following theorem. The interval of the conclusion strongly depends on the lower bound σ_n for the conjunction $P(\bigwedge_{i=0}^n E_i|E_0)$.

THEOREM 9. (Gilio [6]) *If $P(E_i|E_0) = \alpha_i$, for $i = 1, \dots, n$, and $P\left(H \mid \bigwedge_{i=0}^n E_i\right) = \beta$, then*

$$P(H|E_0) \in [\beta\sigma_n, \beta\sigma_n + 1 - \sigma_n], \text{ with}$$

$$\sigma_n = \max \left\{ 0, \sum_{i=1}^n \alpha_i - (n - 1) \right\}.$$

REMARK 10. If P is a Kolmogorov probability, then the bounds are the same for $\sum_{i=1}^n \alpha_i - (n - 1) > 0$. However, if $\sum_{i=1}^n \alpha_i - (n - 1) \leq 0$ and $0 < \beta < 1$, then the interval for $P(H|E_0)$ is the open interval $(0, 1)$. The value 0 (resp. 1) would require that $P\left(\bigwedge_{i=1}^n E_i|E_0\right) = 0$ as the following equation shows

$$P(H|E_0) = P\left(H \mid \bigwedge_{i=0}^n E_i\right) P\left(\bigwedge_{i=1}^n E_i|E_0\right) + P\left(H \mid \neg \bigwedge_{i=1}^n E_i\right) P\left(\neg \bigwedge_{i=1}^n E_i|E_0\right).$$

Therefore, $P\left(\bigwedge_{i=0}^n E_i\right) = P\left(\bigwedge_{i=1}^n E_i|E_0\right)P(E_0) = 0$ and hence $P\left(H \mid \bigwedge_{i=0}^n E_i\right)$ is undefined.

Cut degrades. The width of the interval for $P(H|E_0)$ increases as the number of premises increases. This follows from the facts that its width is $1 - \sigma_n$ and that σ_n is monotonically decreasing. The width of the interval $1 - \sigma_n$ depends on the lower bound of the conjunction, i.e., σ_n . Since this lower bound is zero if $n \geq \sum_{i=1}^n \alpha_i + 1$, the interval for $P(H|E_0)$ is the unit interval if the number of premises is sufficiently high.

3.4. Bayes’ Theorem

Suppose that the prior probability of a certain hypothesis $P(H) = \delta$, the likelihood of the data given both, the hypothesis $H, P(D|H) = \beta$, and the alternative hypothesis $\neg H, P(D|\neg H) = \gamma$, are given. The posterior probability of the hypothesis H given the data D is obtained, if $P(D) > 0$, by Bayes’ Theorem $P(H|D) = \frac{\beta\delta}{\beta\delta + \gamma(1-\delta)}$. The premises of generalized Bayes’ Theorem are $P(H) = \delta, P(E_1|H) = \beta_1, \dots, P(E_n|H) = \beta_n, P(E_1|\neg H) = \gamma_1, \dots, P(E_n|\neg H) = \gamma_n$. In inferential statistics it is often assumed that the E_i ’s are independent and identically distributed. We do neither require conditional independence of the E_i ’s given H nor do we require that $P(E_i|H) = P(E_j|H)$ for $i \neq j$. The conclusion of the generalized Bayes’ Theorem is $P(H|E_1 \wedge \dots \wedge E_n)$.

THEOREM 11. (Wallmann & Kleiter [14], lower bound) *Let $P(H) = \delta$ and let for all $i = 1, \dots, n$, $P(E_i|H) = \beta_i$ and $P(E_i|\neg H) = \gamma_i$. Then:*

- If $\delta \left(\sum_{i=1}^n \beta_i - (n - 1) \right) > 0$, then

$$P(H|E_1 \wedge \dots \wedge E_n) \geq \frac{\delta \left(\sum_{i=1}^n \beta_i - (n - 1) \right)}{\delta \left(\sum_{i=1}^n \beta_i - (n - 1) \right) + (1 - \delta) \min\{\gamma_i\}} .$$

- If $\delta \left(\sum_{i=1}^n \beta_i - (n - 1) \right) \leq 0$, then $P(H|E_1 \wedge \dots \wedge E_n) \geq 0$.

THEOREM 12. (Wallmann & Kleiter [14], upper bound) *Let $P(H) = \delta$ and let for all $i = 1, \dots, n$, $P(E_i|H) = \beta_i$ and $P(E_i|\neg H) = \gamma_i$. Then:*

- If $(1 - \delta) \left(\sum_{i=1}^n \gamma_i - (n - 1) \right) > 0$, then

$$P(H|E_1 \wedge \dots \wedge E_n) \leq \frac{\delta \min\{\beta_i\}}{\delta \min\{\beta_i\} + (1 - \delta) \left(\sum_{i=1}^n \gamma_i - (n - 1) \right)} .$$

- If $(1 - \delta) \left(\sum_{i=1}^n \gamma_i - (n - 1) \right) \leq 0$, then $P(H|E_1 \wedge \dots \wedge E_n) \leq 1$.

Bayes' Theorem does not degrade (for a counter-example see [14]). However, if $n \geq \max \left\{ \sum_{i=1}^n \beta_i + 1, \sum_{i=1}^n \gamma_i + 1 \right\}$, then $\sum_{i=1}^n \beta_i - (n - 1) \leq 0$ and $\sum_{i=1}^n \gamma_i - (n - 1) \leq 0$, so that the interval $[0, 1]$ is obtained. There are two special cases in which Bayes' Theorem degrades. First, if identical likelihoods $\beta_i = \beta$ and $\gamma_i = \gamma$ are assumed. Second, if the values $\gamma_i \in [0, 1]$ are not specified.

3.5. Modus Tollens

Modus Tollens is the inference from $\{\neg B, A \Rightarrow B\}$ to the conclusion $\neg A$. The result for probabilistic Modus Tollens with two premises within the Kolmogorov approach has been derived in [13]. Generalized Modus Tollens is given by

THEOREM 13. (Wallmann & Kleiter [14]) *If $P(\neg E_i) = \alpha_i$, for $i = 1, \dots, n$, and if $P(E_1 \wedge E_2 \wedge \dots \wedge E_n|H) = \beta$, then $P(\neg H) \in [\delta', 1]$, with*

$$\delta' = \begin{cases} 1 - \frac{1-\alpha^*}{\beta} & \text{if } \alpha^* + \beta > 1 \\ 1 - \frac{\sum_{i=1}^n \alpha_i}{1-\beta} & \text{if } \alpha^* + \beta \leq 1 \quad \text{and} \quad \sum_{i=1}^n \alpha_i + \beta < 1 \\ 0 & \text{if } \alpha^* + \beta \leq 1 \quad \text{and} \quad \sum_{i=1}^n \alpha_i + \beta \geq 1 \end{cases},$$

where $\alpha^* = \max\{\alpha_i\}$.

REMARK 14. If P is a Kolmogorov probability, then the upper bound 1 is never correct. Since if $P(\neg H) = 1$, then $P(H) = 0$ and consequently $P(E_1 \wedge E_2 \wedge \dots \wedge E_n|H)$ is undefined. Within the coherence approach an assessment of probability 1 to both premises of Modus Tollens is perfectly alright and leads to probability 1 of the conclusion. A Kolmogorov probability such that $P(\neg B) = 1$ and $P(B|A) = 1$, however, does not exist. If $P(\neg B) = 1$, then $P(A) = 0$, and hence $P(B|A)$ is undefined—a contradiction.

Modus Tollens is special because if $\alpha^* + \beta > 1$, then the interval of its conclusion does not depend on the number of premises n . However, if $\alpha^* + \beta \leq 1$, then it does depend on n . Modus Tollens does not degrade. Moreover, contrary to the other inferences considered so far, the unit interval is not necessarily obtained if the number of premises is large.

3.5.1. Exclusive-Or. SYSTEM O is weaker than SYSTEM P [8,9]. It contains weaker forms of the rules And and Or, Weak-And (Wand) and Weak-Or (Wor). Wor is SYSTEM O equivalent with the rule Exclusive-Or (Xor).

Exclusive-Or is the following rule

$$\text{If } A \Rightarrow C, B \Rightarrow C, \text{ and } \models \neg(A \wedge B), \text{ then } A \vee B \Rightarrow C.$$

The generalized probabilistic version is given by

THEOREM 15. *If $P(H|E_i) = \alpha_i$, for $i = 1, \dots, n$ and $\models \neg(E_i \wedge E_j)$, for $1 \leq i < j \leq n$, then*

$$P(H|E_1 \vee \dots \vee E_n) \in [\min\{\alpha_i\}, \max\{\alpha_i\}].$$

PROOF. If $\models \neg(E_i \wedge E_j)$, then $P(H|E_1 \vee \dots \vee E_n) =$

$$\begin{aligned} &= P(H \wedge E_1|E_1 \vee \dots \vee E_n) + \dots + P(H \wedge E_n|E_1 \vee \dots \vee E_n) \\ &= P(H|E_1)P(E_1|E_1 \vee \dots \vee E_n) + \dots + P(H|E_n)P(E_n|E_1 \vee \dots \vee E_n). \end{aligned}$$

Setting if $\alpha_i = \min\{\alpha_i\}$ (resp. $\alpha_i = \max\{\alpha_i\}$), $P(E_i|E_1 \vee \dots \vee E_n) = 1$ yields the lower (resp. upper) probability for the conclusion. ■

REMARK 16. In the Kolmogorov approach, if for some i, j $\alpha_i \neq \alpha_j$, we obtain the open interval $(\min\{\alpha_i\}, \max\{\alpha_i\})$. In this case we cannot set for some i $P(E_i|E_1 \vee \dots \vee E_n) = 1$, because then $P(E_j) = 0$ for all $j \neq i$ and consequently $P(H|E_j)$ is undefined.

Xor does degrade. However, the interval $[0, 1]$ is not necessarily obtained after addition of a certain number of premises.

4. Probabilistic Validity of Generalized Inference Forms

The key question of this section is whether a certain generalized inference form satisfies one of the probability preservation properties below. The question can be answered by considering the lower bound of the intervals obtained in Section 3. The Kolmogorov approach and the coherence approach often yield different lower bounds. As a consequence, an inference form may satisfy a preservation property relative to one of the approaches while it does not satisfy it with respect to the other approach.

4.1. Preservation Properties

Adams considered four preservation properties [1, p. 1].

1. *Certainty-preservation* If the premises of an inference form have probability 1, then its conclusion has probability 1.
2. *High probability-preservation* If the premises are highly probable, then the conclusion is highly probable, i.e., for all $\delta > 0$ there exists $\epsilon > 0$ such that: If for every premise A it is $P(A) \geq 1 - \epsilon$, then for the conclusion C it holds that $P(C) \geq \delta$.
3. *Positive probability-preservation* If the premises have positive probability, then the conclusion has positive probability.
4. *Minimum probability-preservation* The probability of the conclusion is at least as high as the minimum of the probabilities of the premises. Or equivalently: For all thresholds r : If the probability of each premise is at least r , then the probability of the conclusion is at least r .

The preservation properties above are ordered by strictness. The chain of implications $4 \Rightarrow 3 \Rightarrow 2 \Rightarrow 1$ holds. However, neither of the converse implications is true.

Consider for example Modus Ponens. If $P(B|A) = \beta$ and $P(A) = \alpha$, then the interval for $P(B)$ is $[\alpha\beta, \alpha\beta + 1 - \alpha]$. Modus Ponens $\{A \Rightarrow B, A\} \therefore B$ is consequently

1. Certainty preserving: If $\alpha = 1$ and $\beta = 1$, then $\alpha\beta = 1$.
2. High probability preserving: For $P(H) \geq \delta$, choose $1 - \epsilon = \sqrt{\delta}$.
3. Positive probability preserving: If $\alpha > 0$ and $\beta > 0$, then $\alpha\beta > 0$.
4. Not minimum preserving: In general, it is not the case that $\alpha\beta \geq \min\{\alpha, \beta\}$.

4.2. Certainty-Preservation and High Probability-Preservation

It is important that, while in the Kolmogorov approach high probability and certainty-preservation differ, they are, given the assumption of p -consistent premises, equivalent in the coherence approach [5, 7]. We call a set of premises $\{A_1, \dots, A_n\}$ p -consistent iff the assessment $P(A_1) = \dots = P(A_n) = 1$ is coherent.

THEOREM 17. *Suppose that $\{A_1, \dots, A_n\}$ is p -consistent. Then the inference from $\{A_1, \dots, A_n\}$ to C is in the coherence approach certainty preserving iff it is high probability preserving.*

REMARK 18. In the coherence approach an inference form has been called SYSTEM P valid iff its premises are p -consistent and it is high probability preserving [5]. Contrary to other approaches, System P validity therefore requires p -consistent premises. We mention three such approaches. Adams [1] works with the default assumption: If $P(A) = 0$, then $P(B|A) = 1$ for all B . Hawthorne uses Popper functions. With respect to Popper functions certainty and high probability-preservation are, even without the assumption of p -consistent premises, equivalent [8]. Hawthorne and Makinson [9] employ Kolmogorov probability functions. In Section 4.4, we discuss the inference form Weak-And. It is SYSTEM P valid with respect to these approaches, but not with respect to the coherence approach.

The inference from B to $A \Rightarrow B$ is, for example, relative to the Kolmogorov approach certainty preserving but not high probability preserving. In the coherence approach this inference form is not high probability preserving and therefore, because $\{B\}$ is p -consistent, not certainty preserving.

The inference forms of Section 3 are certainty preserving relative to the coherence approach. This is immediately obtained by considering the lower bound of their conclusion. Consequently, if their premises are p -consistent,

these inference forms are already known to be high probability preserving in the coherence approach.

To show that an inference form is high probability preserving with respect to the coherence approach, we can alternatively determine for every probability of the conclusion δ a “high” probability $1 - \epsilon$ for the premises, such that this probability assessment guarantees that the probability of the conclusion is at least δ . A suitable ϵ can be determined by considering the intervals given in Section 3. Consider, for example, Modus Tollens. Let $\delta > 0$. In order that $P(C) \geq \delta$, the lower bound of $P(C)$, i.e., $1 - \frac{1 - \alpha^*}{\beta}$, may not be less than δ . Therefore, we solve

$$1 - \frac{1 - (1 - \epsilon)}{1 - \epsilon} \geq \delta$$

for ϵ and obtain $1 - \epsilon \geq \frac{1}{2 - \delta}$. A suitable ϵ for the other inference forms can be determined by the same method. We have

THEOREM 19. *Let P be a coherent conditional probability. All inference forms of Section 3 with p -consistent premises are certainty preserving and (consequently) high probability preserving.*

REMARK 20. Although generalized inference forms remain high probability preserving, degradation has a striking consequence. In order to obtain that the probability of the conclusion is at least δ , ϵ is clearly decreasing with increasing n . Since the lower probability of the conclusion decreases as the number of premises increases, for a high probability of the conclusion a higher probability of the premises is necessary with increasing n .

4.3. Positive Probability-Preservation

For $n \geq 2$ premises high probability preservation differs significantly from positive probability preservation. While all of the generalized inference forms considered in Section 3 are high probability preserving, none of them—with the exception of Xor—is positive probability preserving. As already pointed out, in the case of And, Cautious Monotonicity, Cut, and Bayes’ Theorem the lower bound of the conclusion is zero if the number of premises n is sufficiently high.

Moreover, in contrast to their generalizations, some of the inference forms are positive probability preserving. Cut and Bayes’ Theorem are positive probability preserving for $n = 1$. If the sum of the probabilities of the two premises is different from one, then Modus Tollens is also positive probability preserving.

THEOREM 21. *If P is a coherent conditional probability, then the generalizations of And, Cautious Monotonicity, Cut, Modus Tollens and Bayes' Theorem are not positive probability preserving. Generalized Xor is positive probability preserving.*

The Kolmogorov and the coherence approach validate different inference forms. In the Kolmogorov approach generalized Cut, for instance, is positive probability preserving, while this is not the case in the coherence approach.

4.4. Minimum Probability-Preservation

Positive probability preservation and minimum probability preservation differ. Contrary to positive probability preservation, Cut with two premises is, for example, not minimum probability preserving. SYSTEM O is closely connected with minimum probability preservation (for a description of SYSTEM O see [8,9]). All its inference forms are minimum probability preserving. The converse, however, is not true.

The generalization of Xor is minimum probability preserving.

THEOREM 22. *Let P be a coherent or a Kolmogorov conditional probability. Then generalized Xor is minimum probability preserving.*

The SYSTEM O rule Weak-And (Wand) is given by

$$\text{If } A \wedge \neg B \Rightarrow B \text{ and } A \Rightarrow C, \quad \text{then } A \Rightarrow B \wedge C.$$

Wand is central to SYSTEM O and minimum preserving in the Kolmogorov approach. However, a positive probability assessment to $A \wedge \neg B \Rightarrow B$ is incoherent. Hence, from the point of view of coherence, SYSTEM O is not satisfactory.

THEOREM 23. *For every probability assessment P $P(B|A \wedge \neg B) > 0$ is incoherent. In particular, $\{A \wedge \neg B \Rightarrow B, A \Rightarrow C\}$ is not p -consistent.*

PROOF. If P is a coherent probability assessment, then $P(B|A \wedge \neg B) + P(\neg B|A \wedge \neg B) = 1$. Since $P(\neg B|A \wedge \neg B) = 1$, $P(B|A \wedge \neg B) = 0$. ■

REMARK 24. The premises of Wand are not p -consistent. As a consequence Wand is not SYSTEM P valid in the coherence approach (compare Remark 18). In other approaches p -consistency of the premises is not required for SYSTEM P validity (compare Remark 18). Consequently, since And is SYSTEM P valid, and Wand is a special case of And, Wand is SYSTEM P valid in these approaches.

If $P(B|A \wedge \neg B) > 0$, we can conclude in the Kolmogorov framework, that $P(A \wedge \neg B) = 0$ and hence that $P(B|A \wedge \neg B)$ is undefined. Hence, there is no Kolmogorov probability such that $P(B|A \wedge \neg B) > 0$.

5. Conclusions

We have seen that Cautious Monotonicity, Cut, and Exclusive-Or clearly degrade, and that Bayes' Theorem (with some exceptions) and Modus Tollens do not degrade. Moreover, in all the inference forms considered—with the exception of Modus Tollens and Exclusive-Or—the unit interval is obtained even with a “small” number of premises. Narrow intervals may be considered to be better than wide intervals; a more complete knowledge base may be considered to be better than a truncated one [12]. While in general the number of premises and the precision of the conclusion *may* conflict, in generalized inference forms they often *must* conflict.

Degradation does not conflict with the property of monotonicity, but its consequences for information seeking cannot be ignored. On the one hand, the principle of total evidence leads to the selection of the most “recent” interval based on the most specific information. This yields wide intervals, and in many cases even the non-informative $[0, 1]$ interval. On the other hand, a take-the-best strategy leads to the selection of the tightest interval. The according interval is based on the seemingly most “relevant” information with $n = 1$. Since all additional premises are discarded, it would be counterproductive to seek further information, because it would simply be useless.

Degradation is neither “good” nor “bad”. To solve the conflict between precision and specificity requires to counterbalance (i) the width of an interval, (ii) the amount of information it is based upon, and (iii) the position of the interval. The choice depends on pragmatic conditions. An answer to the question which interval should rationally be selected seems to lie outside the domain of probability theory.

It might be supposed that degradation disappears if further constraints are added to the premises. In many cases stochastic independence, for example, leads to point probabilities of the conclusions. Though often presupposed, independence may be a constraint that is too strong. Exchangeability is a related but much weaker assumption. We have shown that in many generalized inference forms exchangeability does not prevent degradation [15].

In general, degradation does not make generalized inference forms probabilistically invalid. Each of the inference forms considered in this contribution is high probability preserving. As already pointed out, the lower probability of the conclusion is often zero if the number of premises is large. Therefore, all of the inference forms—with exception of Exclusive-Or—are not positive probability preserving.

Acknowledgements. Supported by the Austrian Science Foundation (I 141-G15) and the LogICCC Programme of the European Science Foundation. We thank two anonymous referees for their valuable comments.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- [1] ADAMS, E. W., Four probability-preserving properties of inferences, *Journal of Philosophical Logic* 25:1–24, 1996.
- [2] BIAZZO, V., GILIO, A., and G. SANFILIPPO, Coherent conditional previsions and proper scoring rules, in S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager (eds.), *Advances in Computational Intelligence, IPMU (4)*, vol. 300 of *Communications in Computer and Information Science*, Springer, 2012, pp. 146–156.
- [3] COLETTI, G., and R. SCOZZAFAVA, *Probabilistic Logic in a Coherent Setting*, Kluwer, Dordrecht, 2002.
- [4] DE FINETTI, B., *Theory of Probability*, Wiley, London, 1974.
- [5] GILIO, A., Probabilistic reasoning under coherence in System P, *Annals of Mathematics and Artificial Intelligence* 34(1–3):5–34, 2002.
- [6] GILIO, A., Generalization of inference rules in coherence-based probabilistic default reasoning, *International Journal of Approximate Reasoning* 53:413–434, 2012.
- [7] GILIO, A., and G. SANFILIPPO, Probabilistic entailment in the setting of coherence: The role of quasi conjunction and inclusion relation. *International Journal of Approximate Reasoning* 54(4):513–525, 2013.
- [8] HAWTHORNE, J., On the logic of nonmonotonic conditionals and conditional probabilities, *Journal of Philosophical Logic* 25:185–218, 1996.
- [9] HAWTHORNE, J., and D. MAKINSON, The quantitative/qualitative watershed for rules of uncertain inference, *Studia Logica* 86:247–297, 2007.
- [10] KLEITER, G. D., Ockham’s razor in probability logic, in R. Kruse, M. R. Berthold, C. Moewes, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz (eds.), *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, *Advances in Intelligent Systems and Computation*, 190, Springer, 2012, pp. 409–417.

- [11] KRAUS, S., D. LEHMANN., and M. MAGIDOR, Nonmonotonic reasoning, preferential models and cumulative logics, *Artificial Intelligence* 44:167–207, 1990.
- [12] KYBURG, H., and C. M. TENG, *Uncertain Inference*, Cambridge University Press, Cambridge, 2001.
- [13] WAGNER, C. G., Modus tollens probabilized, *British Journal for the Philosophy of Science* 55:747–753, 2004.
- [14] WALLMANN, C., and G. D. KLEITER, Beware of too much information, in T. Kroupa, and J. Vejnarova (eds.), *Proceedings of the 9th Workshop on Uncertainty Processing, WUPES*, Faculty of Managment, University of Economics, Prague, 2012, pp. 214–225.
- [15] WALLMANN, C., and G. D. KLEITER, Exchangeability in probability logic, in S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager (eds.), *IPMU (4)*, vol. 300 of *Communications in Computer and Information Science*, Springer, 2012, pp. 157–167.

C. WALLMANN, G. D. KLEITER
Department of Psychology
University of Salzburg
Hellbrunnerstr. 34
Salzburg, Austria
christian.wallmann@stud.sbg.ac.at

G. D. KLEITER
gernot.kleiter@sbg.ac.at