On the Symbiosis of Two Concepts of Conditional Interval Probability*

KURT WEICHSELBERGER
Ludwig-Maximilians-University Munich, Germany

THOMAS AUGUSTIN
Ludwig-Maximilians-University Munich, Germany

Abstract

This paper argues in favor of the thesis that two different concepts of conditional interval probability are needed, in order to serve the huge variety of tasks conditional probability has in the classical setting of precise probabilities. We compare the commonly used intuitive concept of conditional interval probability with the canonical concept, and see, in particular, that the canonical concept is the appropriate one to generalize the idea of transition kernels to interval probability: only the canonical concept allows reconstruction of the original interval probability from the marginals and conditionals, as well as the powerful formulation of Bayes Theorem.

Keywords

conditional interval probability, intuitive concept of conditional interval probability, canonical concept of conditional interval probability, conditioning, updating, theorem of total probability, Markov chains, Bayes theorem, decision theory

1 Introduction

In the last years a comprehensive theory of interval probability has been developed which systematically generalizes Kolmogorov's axiomatic approach to classical probability. Just as in Kolmogorov's approach, the basic axioms have to be supplemented by appropriate concepts of independence and by a definition of conditional probability.

The goal of the theory of interval probability is not only the creation of methods for dealing with imprecise probability but also a systematic one: the establishment of a body of definitions and results comparable to the analogous elements of

*We thank Anton Wallner for his valuable remarks and his support.

the classical theory with respect to rigidity and efficiency but with a much wider field of appropriate application.

While the system of axioms describing the properties of probability assignments is thoroughly discussed in [27] (see also [25] and [26]), the necessary supplements concerning independence and conditional probability are not included in that volume. A report summarizing basic aspects results in the statement that there is need for two different definitions of conditional probability associated with different roles in employing interval probability: the intuitive concept of conditional probability and the canonical concept of conditional probability ([26]).

The intuitive concept of conditional probability is widely used as the only generalization of classical conditional probability to imprecise probability in general (for a recent study in the context of numerical possibility theory, see [8], section 6). This way of generalizing conditional probability was rigorously justified by Walley [22], who derived it from coherence considerations between gambles and contingent gambles. It is almost exclusively used in statistical inference with imprecise probabilities: In particular, it underlies Walley's imprecise Dirichlet model (cf. [23], see also, e.g., [3] and [31]), and it is often understood as self-evident in robust Bayesian inference (e.g. [24], [18]).

Mainly in the area of artificial intelligence, often another definition of conditional interval probability is applied. It dates back to Dempster [10] and his proposed method of statistical inference. Since Shafer [19] it is often used isolated from its original motivation as *Dempster's rule of conditioning*. It has experienced many modifications, see [30] for a comparison of different proposals.

Only very few authors have argued in favor of a symbiosis of different concepts of conditional probabilities. Dubois and Prade [11] use the intuitive concept for what they call 'focusing', and Dempster's rule for 'conditioning'. Halpern and Fagin [12] stress that there are different ways to understand belief functions, a fact which naturally leads to different concepts of conditional probability.

Weichselberger argues that the intuitive concept has to be supplemented by the canonical concept, which in rare situations produces the same result as the concept proposed by Dempster. Although in many situations the canonical concept does not qualify for being interpreted as an assignment of interval probability itself, it serves as the inevitable bearer of information for solving important problems. This is not surprising, since even in classical theory conditional probability has two different roles: sometimes as an information of its own value, but in other cases as a tool allowing the derivation of other quantities. In the theory of interval probability canonical conditional probability and canonical conditional expectation can be used for such purposes, irrespective of their qualification as interval probability or as interval expectation. The relation between the two concepts of conditional interval probability with respect to different situations is the subject of the present article. After introducing basic requirements in Section 2, in Sections 3 to 6 we compare the consequences of the employment of each concept with respect to some relevant aspect of conditioning. Section 7 contains the conclusions which can be drawn.

2 Basic Concepts

Every probability measure in the usual sense, i.e. every set function $p(.)$ satisfying Kolmogorov's axioms, is called a *classical probability*. The set of all classical probabilities on a measurable space $(\Omega;\mathcal{A})$ will be denoted by $\mathcal{K}(\Omega;\mathcal{A})$. According to [27] axioms for interval probability $P(.) = [L(.),U(.)]$ can be obtained by describing the relation between the non-additive set functions $L(.)$ and $U(.)$ and the set of classical probabilities being in accordance with them. Set functions $P(.)\colon \mathcal{A} \rightarrow \mathcal{Z}_0 := \{[L;U] \mid 0 \leq L \leq U \leq 1\}, A \mapsto P(A) = [L(A);U(A)]$, with $\mathcal{M} := \{p(.) \in \mathcal{K}(\Omega;\mathcal{A}) \mid L(A) \leq p(A) \leq U(A), \forall A \in \mathcal{A}\} \neq \emptyset$ are called *R-probability* with *structure* $\mathcal{M}$. If additionally $\inf_{p(.)\in\mathcal{M}} p(A) = L(A)$, and $\sup_{p(.)\in\mathcal{M}} p(A) = U(A)$, $\forall A \in \mathcal{A}$, hold, then $P(.)$ is *F-probability*. (With allowance to the different attitudes of Kolmogorov and de Finetti towards σ-additivity R-probability materially corresponds to a probability assignment 'avoiding sure loss' described by interval limits and F-probability to a 'coherent' assignment by interval limits.) The triple $\mathcal{F} = (\Omega;\mathcal{A};L(.))$ is called an *F-probability field.*

A non-empty subset $\mathcal{V}$ of $\mathcal{M}$ is called a *prestructure* of $\mathcal{F} = (\Omega;\mathcal{A};L(.))$ if the following equations hold: $\inf_{p(.)\in\mathcal{V}} p(A) = L(A)$, $\sup_{p(.)\in\mathcal{V}} p(A) = U(A)$, $\forall A \in \mathcal{A}$. The concept of independence[1] is introduced by

Definition 1 *Let $\mathcal{F} = (\Omega;\mathcal{A};L(.))$ be an F-probability field with structure $\mathcal{M}$ and let $\mathcal{C}_i$, $i = 1,2$, be partitions of Ω. Then $\mathcal{C}_1$ and $\mathcal{C}_2$ are* mutually independent, *if the set $\mathcal{M}_I = \{p(.) \in \mathcal{M} \mid p(A_1 \cap A_2) = p(A_1)\cdot p(A_2), \forall A_i \in \mathcal{C}_i, i = 1,2\}$ serves as a prestructure of the field $\mathcal{F}$.* □

In [26] this definition is illustrated in the case of a fourfold-table. As also mentioned there, apart from cases for which at least one of the marginal probabilities is a classical probability, the structure $\mathcal{M}$ will always contain classical probabilities with some dependence of $\mathcal{C}_1$ and $\mathcal{C}_2$ in the classical sense.

The classical concept of conditional probability can be generalized to interval probability in two different ways, generating on the one hand the intuitive concept, on the other hand the canonical concept of conditional probability, two concepts with different properties in many respects.

[1] The question how to generalize the notion of independence has received considerable attention (see, e.g., [4] for a survey and [13] for a comprehensive treatment in the context of random sets). Recently, in particular the concepts of epistemic irrelevance and independence, introduced by Walley [22], have been investigated in detail (see, among others, [7], [14], [16], [17], [21], [20]).

In the context studied here the most natural definition is to call two partitions independent if the structure of the underlying F-probability field is generated by the set of independent classical probabilities (cf. [25], [26]). This way of defining independence corresponds to the notion of strong extension ([4], [6]).

Definition 2 *Let $\mathcal{F} = (\Omega; \mathcal{A}; L(.))$ be an F-probability field with structure $\mathcal{M}$, and $\mathcal{C}$ be a partition of Ω where $U(C) > 0$, $\forall C \in \mathcal{C}$. With $\mathcal{M}_C := \{p(.) \in \mathcal{M} \mid p(C) > 0\}$ the* intuitive concept of conditional probability *is given by defining $iP_{\mathcal{C}}(A \mid C) = [iL_{\mathcal{C}}(A \mid C); iU_{\mathcal{C}}(A \mid C)]$, where*

$$iL_{\mathcal{C}}(A \mid C) = \inf_{p(.)\in\mathcal{M}_C} \frac{p(A \cap C)}{p(C)},\ iU_{\mathcal{C}}(A \mid C) = \sup_{p(.)\in\mathcal{M}_C} \frac{p(A \cap C)}{p(C)}, \forall A \in \mathcal{A}, \forall C \in \mathcal{C}.$$

□

It can be demonstrated that this definition generates a conditional F-field for every C with $U(C) > 0$. The motivation of employing the intuitive concept is straightforward: As long as $L(C) > 0$ it may be understood as the transition from the structure $\mathcal{M}$ to the structure $i\mathcal{M}_{\mathcal{C}}(. \mid C) = \{p(. \mid C) \mid p(.) \in \mathcal{M}\}$, which consists exactly of all classical conditional probabilities corresponding to elements of the original structure $\mathcal{M}$. The conditional interval expectation of any gain function $G(.)$ — defined for all elements E of Ω — therefore is calculated as $i\mathbb{E}(G(.) \mid C) = [i\mathbb{L}(G(.) \mid C); i\mathbb{U}(G(.) \mid C)] = \{E_p(G(.)) \mid p(.) \in i\mathcal{M}_{\mathcal{C}}(. \mid C)\}$.

Weichselberger ([25], [26]) argues that the intuitive concept has to be supplemented by a concept, which is derived from a canon of desirable properties, and therefore is called the *canonical concept of conditional interval probability*.

Definition 3 *Let $\mathcal{F} = (\Omega; \mathcal{A}; L(.))$ be an F-probability field and $\mathcal{C}$ be a partition of Ω where $L(C) > 0$, $\forall C \in \mathcal{C}$. The canonical concept of conditional probability is given by $L_{\mathcal{C}}(A \mid C) := \frac{L(A\cap C)}{L(C)}$ and $U_{\mathcal{C}}(A \mid C) := \frac{U(A\cap C)}{U(C)}$, $\forall A \in \mathcal{A}$, $\forall C \in \mathcal{C}$.*

The canonical concept of conditional expectation of the gain function $G(.)$ for each $C \in \mathcal{C}$ with $L(C) > 0$ is defined as $\mathbb{E}[G(.) \mid C] := [\mathbb{L}(G(.) \mid C); \mathbb{U}(G(.) \mid C)]$ with $\mathbb{L}(G(.) \mid C) := \frac{\mathbb{L}(G(.)\cap C)}{L(C)}$, $\mathbb{U}(G(.) \mid C) := \frac{\mathbb{U}(G(.)\cap C)}{U(C)}$ and $\mathbb{L}(G(.) \cap C) = \inf_{p\in\mathcal{M}} \sum_{E\subseteq A} G(E) \cdot p(E)$, $\mathbb{U}(G(.) \cap C) = \sup_{p\in\mathcal{M}} \sum_{E\subseteq A} G(E) \cdot p(E)$. □

Three simple examples demonstrate the different types of conditional probability according to the canonical concept. Each is constructed from an F-probability field on $\Omega = E_1 \cup E_2 \cup E_3$ with $\mathcal{A} = \mathcal{P}(\Omega)$, and the same partition $\mathcal{C} = (C_1, C_2)$ with $C_1 = E_1 \cup E_2$, $C_2 = E_3$ is considered. Example 1 describes a constellation $(\mathcal{F}; \mathcal{C})$ with conditional F-probability according to the canonical concept.

Example 1 *An F-probability field $\mathcal{F}^{(1)}$ is given by*

$$\begin{aligned} P(E_1) &= [0.10; 0.30] & P(E_1 \cup E_2) &= [0.40; 0.60] \\ P(E_2) &= [0.20; 0.45] & P(E_1 \cup E_3) &= [0.55; 0.80] \\ P(E_3) &= [0.40; 0.60] & P(E_2 \cup E_3) &= [0.70; 0.90] \end{aligned}$$

Because of $L(C_1) = 0.40$, $U(C_1) = 0.60$ the conditional probability according to *the canonical concept is given by*

$$\begin{array}{ll} P_{\mathcal{C}}(E_1 \mid C_1) = [0.25; 0.50] & P_{\mathcal{C}}(E_1 \mid C_2) = [0] \\ P_{\mathcal{C}}(E_2 \mid C_1) = [0.50; 0.75] & P_{\mathcal{C}}(E_2 \mid C_2) = [0] \\ P_{\mathcal{C}}(E_3 \mid C_1) = [0] & P_{\mathcal{C}}(E_3 \mid C_2) = [1] \end{array}$$

It is easily seen that in this case both conditional probability fields are F-probability. In addition the results may be compared with those from applying the intuitive concept:

$$\begin{array}{ll} iP_{\mathcal{C}}(E_1 \mid C_1) = [0.182; 0.600] & iP_{\mathcal{C}}(E_1 \mid C_2) = [0] \\ iP_{\mathcal{C}}(E_2 \mid C_1) = [0.400; 0.818] & iP_{\mathcal{C}}(E_2 \mid C_2) = [0] \\ iP_{\mathcal{C}}(E_3 \mid C_1) = [0] & iP_{\mathcal{C}}(E_3 \mid C_2) = [1] \end{array}$$

□

It can be shown that interval limits resulting from the intuitive concept cannot be narrower than those arising from the canonical one (cf., e.g., [22], p. 301). In general for all $C \in \mathcal{C}$, $iP_{\mathcal{C}}(E_i \mid C) \supsetneqq P_{\mathcal{C}}(E_i \mid C)$ holds. Both concepts coincide if the marginals consist of classical probabilities.[2] Example 2 shows a constellation $(\mathcal{F}; \mathcal{C})$ for which the conditional probability according to the canonical concept possesses R-quality but not F-quality.

Example 2 *The F-field $\mathcal{F}^{(2)}$ is given by*

$$\begin{array}{ll} P(E_1) = [0.10; 0.25] & P(E_1 \cup E_2) = [0.40; 0.60] \\ P(E_2) = [0.20; 0.40] & P(E_1 \cup E_3) = [0.60; 0.80] \\ P(E_3) = [0.40; 0.60] & P(E_2 \cup E_3) = [0.75; 0.90] \end{array}$$

The conditional probability according to the canonical concept now reads as follows:

$$\begin{array}{ll} P_{\mathcal{C}}(E_1 \mid C_1) = [0.250; 0.417] & P_{\mathcal{C}}(E_1 \mid C_2) = [0] \\ P_{\mathcal{C}}(E_2 \mid C_1) = [0.500; 0.667] & P_{\mathcal{C}}(E_2 \mid C_2) = [0] \\ P_{\mathcal{C}}(E_3 \mid C_1) = [0] & P_{\mathcal{C}}(E_3 \mid C_2) = [1] \end{array}$$

The fact, that $L_{\mathcal{C}}(E_1 \mid C_1) + U_{\mathcal{C}}(E_2 \mid C_1) \neq 1$ and $L_{\mathcal{C}}(E_2 \mid C_1) + U_{\mathcal{C}}(E_1 \mid C_1) \neq 1$ makes it clear, that $P_{\mathcal{C}}(. \mid C_1)$ is not F-probability. On the other hand the assignment $p_{\mathcal{C}}(E_1 \mid C_1) = 0.4$, $p_{\mathcal{C}}(E_2 \mid C_1) = 0.6$, $p_{\mathcal{C}}(E_3 \mid C_1) = 0.0$ is an element of the structure of this field: The canonical concept here produces an R-field, but not an F-field. Again the intuitive concept produces an F-probability-field with wider interval limits: $iP_{\mathcal{C}}(E_1 \mid C_1) = [0.200; 0.556]$ and $iP_{\mathcal{C}}(E_2 \mid C_1) = [0.444; 0.800]$ completed by the same trivial interval limits as in Example 1. □

Example 3 describes a constellation for which canonical conditional probability has not even R-quality, since it contains intervals for which $L > U$ holds.

[2] For an attractive example for this special situation see the nonparametric predictive inference discussed in [2].

Example 3 *The F-field $\mathcal{F}^{(3)}$ is given through*

$$\begin{array}{ll} P(E_1) = [0.16; 0.18] & P(E_1 \cup E_2) = [0.40; 0.60] \\ P(E_2) = [0.22; 0.42] & P(E_1 \cup E_3) = [0.58; 0.78] \\ P(E_3) = [0.40; 0.60] & P(E_2 \cup E_3) = [0.82; 0.84] \end{array}$$

The canonical concept produces: $P_C(E_1 \mid C_1) = [0.40; 0.30]$ and $P_C(E_2 \mid C_1) = [0.55; 0.70]$ and the same trivial interval limits as the foregoing examples. Since $L_C(E_1 \mid C_1) > U_C(E_1 \mid C_1)$, it is impossible to find K-functions in accordance with the interval limits: A structure does not exist. Concerning the intuitive concept, there are no problems: $iP_C(E_1 \mid C_1) = [0.276; 0.450]$ and $iP_C(E_2 \mid C_1) = [0.550; 0.724]$. □

It is obvious that in a case like this the outcome of the canonical concept cannot be interpreted as interval probability in the usual sense. In order to allow the employment of the word *probability*, the usage of this expression has to be extended.

Definition 4 *Given a sample space Ω and a σ-field $\mathcal{A}$ of random events in Ω, $P(A) = [L(A); U(A)]$, $\forall A \in \mathcal{A}$, is named* O-probability*, if $0 \leq L(A), U(A) \leq 1$, $\forall A \in \mathcal{A}$.* □

$P(A), A \in \mathcal{A}$, need not be intervals, $L(A)$ may be larger than $U(A)$. It will be shown in the following sections that the canonical concept produces results which are bearers of important information, even they do not qualify for being interpreted as R-probability or as interval expectation.

3 Independence and Conditional Probability

In the classical theory mutual independence of two partitions $\mathcal{C}_1$ and $\mathcal{C}_2$ can be characterized by

$$p(A_1 \mid A_2) = p(A_1), \quad \forall A_1 \in \mathcal{C}_1, \forall A_2 \in \mathcal{C}_2 : p(A_2) \neq 0. \tag{1}$$

If the intuitive concept of conditional interval probability and the definition of independence along the lines of Definition 1 are applied, this appealing property does not hold in general. This fact led to the introduction of the notions of epistemic irrelevance and epistemic independence (see the references in footnote 1), which use variants of (1) to define independence.

In contrast, (1) extends to interval probability if the canonical concept of conditional probability is employed (and $L(A_2) \neq 0$). According to Definition 1 $L(A_1 \cap A_2) = L(A_1) \cdot L(A_2)$ and $U(A_1 \cap A_2) = U(A_1) \cdot U(A_2)$ hold. This leads to

Theorem 1 *If $\mathcal{F} = (\Omega; \mathcal{A}; L(.))$ is an F-probability field and $\mathcal{C}_1$, $\mathcal{C}_2 \subseteq \mathcal{A}$ are partitions of Ω, the statements a) and b) are equivalent:*

1. $P_{\mathcal{C}_2}(A_1 \mid A_2) = P(A_1), \quad \forall A_1 \in \mathcal{C}_1, A_2 \in \mathcal{C}_2 : L(A_2) \neq 0.$

2. *$\mathcal{C}_1$ and $\mathcal{C}_2$ are mutually independent.* □

It is, therefore, in this case guaranteed that the canonical concept produces conditional F-probability fields. On the other hand: Since the interval limits of the intuitive concept are generally wider than that of the canonical one, in the case of mutual independence $iP_{\mathcal{C}_2}(A_1|A_2) \supsetneqq P(A_1)$ must be expected. Employing the model of double-dichotomy this phenomenon is demonstrated in Example 4.

Example 4 *Let $\mathcal{F} = (\Omega; \mathcal{A}; L(.))$ be an F-probability field with $\Omega_4 = E_1 \cup E_2 \cup E_3 \cup E_4$, $\mathcal{A} = \mathcal{P}(\Omega_4)$ and*

$$\begin{array}{ll} P(E_1) = [0.08; 0.21] & P(E_1 \cup E_2) = [0.20; 0.30] \\ P(E_2) = [0.06; 0.18] & P(E_1 \cup E_3) = [0.40; 0.70] \\ P(E_3) = [0.28; 0.49] & P(E_1 \cup E_4) = [0.33; 0.66] \\ P(E_4) = [0.21; 0.48] & \end{array}$$

(The remaining components of this F-field follow from $L(A) + U(\neg A) = 1, \forall A \in \mathcal{A}$.) Let two partitions be given by $\mathcal{C}_1 = (A_1, \neg A_1)$, where $A_1 = (E_1 \cup E_2)$, and $\mathcal{C}_2 = (A_2, \neg A_2)$, where $A_2 = (E_1 \cup E_3)$. By means of a four-fold table independence of $\mathcal{C}_1$ and $\mathcal{C}_2$ is directly controlled:

$P(E_1)$=[0.08; 0.21]	$P(E_1)$=[0.06; 0.18]	$P(E_1 \cup E_2)$=[0.20; 0.30]
$P(E_3)$=[0.28; 0.49]	$P(E_4)$=[0.21; 0.48]	$P(E_3 \cup E_4)$=[0.70; 0.80]
$P(E_1 \cup E_3)$=[0.40; 0.70]	$P(E_2 \cup E_4)$=[0.30; 0.60]	$P(\Omega_4)$=[1]

$$L(E_1 \cup E_4) = \max(L(E_1) + L(E_4), 1 - U(E_2) - U(E_3)) = 0.33$$
$$U(E_1 \cup E_4) = \min(U(E_1) + U(E_4), 1 - L(E_2) - L(E_3)) = 0.66.$$

The canonical concept of conditional probability produces:

$$L_{\mathcal{C}_2}(A_1 \mid A_2) = L_{\mathcal{C}_2}(E_1 \cup E_2 \mid E_1 \cup E_3) = \frac{L(E_1)}{L(E_1 \cup E_3)} = \frac{0.08}{0.40} = 0.20 = L(E_1 \cup E_2)$$
$$U_{\mathcal{C}_2}(A_1 \mid A_2) = U_{\mathcal{C}_2}(E_1 \cup E_2 \mid E_1 \cup E_3) = \frac{U(E_1)}{U(E_1 \cup E_3)} = \frac{0.21}{0.70} = 0.30 = U(E_1 \cup E_2).$$

The intuitive concept leads to

$$iL_{\mathcal{C}_2}(A_1 \mid A_2) = \inf_{\mathcal{M}} p(E_1 \mid E_1 \cup E_3) = \frac{0.08}{0.08 + 0.49} = 0.140 < 0.20$$
$$iU_{\mathcal{C}_2}(A_1 \mid A_2) = \sup_{\mathcal{M}} p(E_1 \mid E_1 \cup E_3) = \frac{0.21}{0.21 + 0.28} = 0.429 > 0.30.$$ □

The conclusion from these results is evident: If it is of importance, that in the case of mutual independence conditional and marginal probability are equal, then the canonical concept of conditional probability must be employed.

4 Updating with Conditional Probability

The essential aspects concerning updating by means of conditional interval probability already become clear in the simple case of two states, A_1 and A_2, and two (or later, three) possible diagnoses, B_1 and B_2 (and later, B_3). If the overall probability is given by an F-field $\mathcal{F} = (\Omega; \mathcal{P}(\Omega); L(.))$ with $|\Omega| = 4$, one has

$P(A_1 \cap B_1) = [L_{11}; U_{11}]$	$P(A_1 \cap B_2) = [L_{12}; U_{12}]$	$P(A_1) = [L_{1.}; U_{1.}]$
$P(A_2 \cap B_1) = [L_{21}; U_{21}]$	$P(A_2 \cap B_2) = [L_{22}; U_{22}]$	$P(A_2) = [L_{2.}; U_{2.}]$
$P(B_1) = [L_{.1}; U_{.1}]$	$P(B_2) = [L_{.2}; U_{.2}]$	$P(\Omega_4) = [1]$

While the prior probability of state A_1 is given by $P(A_1)$, updating in case of diagnosis B_1 produces the conditional probability of $(A_1 \cap B_1)$ given B_1. If more than two diagnoses are possible, it is important to ensure that the process of updating is associative: Does stepwise learning lead to the same result as instantaneous learning? In the case of classical probability the answer is affirmative.

An F-probability field $\mathcal{F} = (\Omega; \mathcal{P}(\Omega); L(.))$ with $|\Omega| = 6$ is given by:

P_{11}	P_{12}	P_{1S}	P_{13}	$P_{1.}$
P_{21}	P_{22}	P_{2S}	P_{23}	$P_{2.}$
$P_{.1}$	$P_{.2}$	$P_{.S}$	$P_{.3}$	$[1]$

where

$$P_{ij} := P(A_i \cap B_j)$$
$$P_{i.} := P(A_i),\ i = 1, 2$$
$$P_{.j} := P(B_j),\ j = 1, 2, 3$$
$$P_{iS} := P(A_i \cap (B_1 \cup B_2))$$
$$P_{.S} := P(B_1 \cup B_2)$$

and in an analogous way for L and U.

Let instantaneous learning immediately transfer the information from Ω to B_1, while stepwise learning leads from Ω to $B_1 \cup B_2$ and from there to B_1. A method of updating can only be accepted, if the final result is equal in both cases.

For the intuitive concept it is sufficient to remember that for classical probability the equation $p(A \mid B_1) = \frac{p(A \cap B_1 \mid B_1 \cup B_2)}{p(B_1 \mid B_1 \cup B_2)}$ is valid. This is not only the reason, why associativity holds for updating with the classical conditional probability; it also means that $i\mathcal{M}(. \mid B_1) = \left\{p(. \mid B_1) \mid p(.) \in i\mathcal{M}(. \mid B_1 \cup B_2)\right\}$ must be true and updating with the intuitive concept of conditional probability produces the same results for instantaneous and for stepwise learning.

With respect to the canonical concept the first step of information ("$B_1 \cup B_2$") produces the conditional probability field with the following interval limits:

$\left[\frac{L_{11}}{L_{.S}}; \frac{U_{11}}{U_{.S}}\right]$	$\left[\frac{L_{12}}{L_{.S}}; \frac{U_{12}}{U_{.S}}\right]$	$\left[\frac{L_{1S}}{L_{.S}}; \frac{U_{1S}}{U_{.S}}\right]$
$\left[\frac{L_{21}}{L_{.S}}; \frac{U_{21}}{U_{.S}}\right]$	$\left[\frac{L_{22}}{L_{.S}}; \frac{U_{22}}{U_{.S}}\right]$	$\left[\frac{L_{2S}}{L_{.S}}; \frac{U_{2S}}{U_{.S}}\right]$
$\left[\frac{L_{.1}}{L_{.S}}; \frac{U_{.1}}{U_{.S}}\right]$	$\left[\frac{L_{.2}}{L_{.S}}; \frac{U_{.2}}{U_{.S}}\right]$	$[1]$

The second step of information ("B_1") leads to the interval limits $\frac{L_{11}}{L_{.S}} : \frac{L_{.1}}{L_{.S}} = \frac{L_{11}}{L_{.1}}$, $\frac{U_{11}}{U_{.S}} : \frac{U_{.1}}{U_{.S}} = \frac{U_{11}}{U_{.1}}$, which are the same as if the information "B_1" had been given at once. Therefore the canonical concept satisfies the necessary condition for reasonable updating as well.

It may be concluded that in principle each of the two concepts can be employed for updating. Since the intuitive concept guarantees the F-property of the outcome it should be preferred under usual circumstances.

5 Transfer of Information

The idea of conditional probability often is employed in designing new models, combining marginal probability derived from one source of information, with conditional probability gained from another source. In particular the theory of Markov chains relies on this principle: The dynamic evolution is completely described by specifying an initial distribution and a matrix of transition probabilities, consisting of the conditional probabilities to reach a state i given state j.

A necessary condition for the qualification of any concept of conditional probability with respect to such transfer obviously is the possibility to reconstruct an F-probability field by means of marginal probability and conditional probability. It was demonstrated in [26], that this reconstruction need not be possible if the intuitive concept is employed: different F-fields may be equal with respect to marginal probability and to intuitive conditional probability for a certain partition. This phenomenon is quite common for the intuitive concept: There are very rare borderline cases where it is possible to determine an F-field uniquely by means of marginal probability and the respective intuitive conditional probability.

On the other hand, reconstruction of an F-probability field using the marginal probability of a partition together with the canonical conditional probability is practicable, if a so called *laminar constellation* in the following sense is given.

Definition 5 *i)* $(\mathcal{A}_L, \mathcal{A}_U)$ *is named a* support *of the F-field* $\mathcal{F} = (\Omega; \mathcal{A}; L(.))$ *with structure* $\mathcal{M}$*, if the set of equations:* $L(A) \leq p(A), \forall A \in \mathcal{A}_L$*, and* $p(A) \leq U(A), \forall A \in \mathcal{A}_U$*, is sufficient to determine* $\mathcal{M}$*.*

ii) A constellation $(\mathcal{F}, \mathcal{C})$*, consisting of an F-field* $\mathcal{F} = (\Omega; \mathcal{A}; L(.))$ *and a partition* $\mathcal{C}$ *of* Ω*, is named a* laminar constellation*, if there exists a support* $(\mathcal{A}_L, \mathcal{A}_U)$ *of* $\mathcal{F}$*, so that for each* $A \in \mathcal{A}_L \cup \mathcal{A}_U$ *one of the two following conditions is satisfied:*

1. $\exists C_{(1)}, \ldots, C_{(q)} \in \mathcal{C} : A = \bigcup_{i=1}^{q} C_{(i)}$.
2. $\exists C \in \mathcal{C} : A \subset C$. □

This definition characterizes constellations, where all information about the structure $\mathcal{M}$ — and therefore about $\mathcal{F}$ itself — is contained only in the marginal probability on $\mathcal{C}$ or in events which are subsets of single elements of the partition.

If laminarity of the constellation is given, reconstruction of the original F-field by means of marginal probability of $\mathcal{C}$ and the canonical conditional probabilities for all $C \in \mathcal{C}$ is possible, irrespective of the quality of the canonical conditional probabilities, since for each A satisfying condition a) the interval limits are determined by the marginal probability and for each A satisfying condition b) the interval limits are to be reconstructed by $L(A) = L_{\mathcal{C}}(A \mid C) \cdot L(C) = \frac{L(A)}{L(C)} \cdot L(C)$ and $U(A) = U_{\mathcal{C}}(A \mid C) \cdot U(C) = \frac{U(A)}{U(C)} \cdot U(C)$.

The reconstruction of an F-field using conditional O-probability is demonstrated in Example 5 for a sample space of size 3.

Example 5 *Let an F-field $\mathcal{F} = (\Omega_3; \mathcal{P}(\Omega_3); L(.))$ be given by:*

$$P(E_1) = [0.16; 0.21] \quad P(E_2) = [0.22; 0.42] \quad P(E_3) = [0.40; 0.60].$$

The partition $\mathcal{C} = (C_1, C_2)$ with $C_1 = E_1 \cup E_2$ and $C_2 = E_3$ leads to $P(C_1) = [0.40; 0.60]$, $P(C_2) = [0.40; 0.60]$. It is obvious, that this is a laminar constellation: E_1 and E_2 obey condition b), E_3 satisfies condition a). The interval limits of conditional probability according to the canonical concept are:

$$\begin{array}{ll} L_{\mathcal{C}}(E_1 \mid C_1) = 0.40 & U_{\mathcal{C}}(E_1 \mid C_1) = 0.35 \\ L_{\mathcal{C}}(E_2 \mid C_1) = 0.55 & U_{\mathcal{C}}(E_2 \mid C_1) = 0.70 \\ L_{\mathcal{C}}(E_3 \mid C_2) = 1 & U_{\mathcal{C}}(E_3 \mid C_2) = 1\,. \end{array}$$

Therefore $P_{\mathcal{C}}(E_1 \mid C_1) = [0.40; 0.35]$, $P_{\mathcal{C}}(E_2 \mid C_1) = [0.55; 0.70]$ is an assignment which can not be interpreted as a generalization of classical probability, but it is useful for reconstructing $\mathcal{F}$:

$$\begin{aligned} L(E_1) &= L_{\mathcal{C}}(E_1 \mid C_1) \cdot L(C_1) = 0.40 \cdot 0.40 = 0.16 \\ U(E_1) &= U_{\mathcal{C}}(E_1 \mid C_1) \cdot U(C_1) = 0.35 \cdot 0.60 = 0.21 \\ L(E_2) &= L_{\mathcal{C}}(E_2 \mid C_1) \cdot L(C_1) = 0.55 \cdot 0.40 = 0.22 \\ U(E_2) &= U_{\mathcal{C}}(E_2 \mid C_1) \cdot U(C_1) = 0.70 \cdot 0.60 = 0.42 \\ L(E_3) &= L_{\mathcal{C}}(E_3 \mid C_2) \cdot L(C_2) = 1 \cdot 0.40 = 0.40 \\ U(E_3) &= U_{\mathcal{C}}(E_3 \mid C_2) \cdot U(C_1) = 1 \cdot 0.60 = 0.60\,. \end{aligned}$$

Because of the laminarity of the constellation $(\mathcal{F}, \mathcal{C})$ these interval limits are sufficient to reconstruct $\mathcal{F}$. □

The Theorem of Total Probability can be formulated as

Corollary 1 *If $(\mathcal{F}; \mathcal{C})$ is a laminar constellation, the F-field $\mathcal{F}$ is uniquely determined by the marginal probability field for $\mathcal{C}$ and by the canonical conditional probability fields resulting for each $C \in \mathcal{C}$, irrespective of the F- or R- or O-quality of the conditional fields.* □

This result allows interpretations:

1. If the conditional probabilities do not possess F-(R-)quality, there *cannot be any set* of conditional F-(R-)probabilities which allows to reconstruct the given F-field through the given marginal F-probability.

2. If the process of matching a given marginal F-probability with given canonical conditionals results in a field not possessing F-(R-)quality, it is *impossible to find* an F-(R-)field with this marginal and with these conditionals.

Therefore, the results of transfer of information from one model to another to some extent can be foreseen:

1. If $P_C(. \mid .)$ describes an F-field, matching with marginal F-probability always produces an F-field.

2. If $P_C(. \mid .)$ describes an R-field, matching with marginal F-probability produces either an F-field or an R-field which does not possess the F-quality.

3. If $P_C(. \mid .)$ does not fit to an R-field, nothing can be predicted about quality of the outcome, if it is matched with marginal F-probability.

6 The Theorem of Bayes

The Theorem of Bayes is an important result of classical probability theory. While it is of highest significance for any subjectivistic school, even the objectivistic view sometimes finds conditions, under which it is legitimate to accept a certain prior information which is described by classical probability. On the other hand even the subjectivist cannot deny that in most practically relevant cases the choice of a particular classical prior is at least highly debatable.

Therefore this is a situation inviting to propose the employment of interval probability. If a successful transfer of the Theorem of Bayes into the theory of interval probability can be achieved, a strong argument favouring the efficiency of this theory is presented.[3] Ambiguity, however, — distinguishing interval and classical probability — does not obey to those laws which are the basis of the Theorem of Bayes in the classical theory. It should therefore not be expected that the roles of this theorem in classical probability and in generalized probability are the same. References to the obvious limitations for the efficiency of particular types of this theorem have been given only recently ([29], [1]).

In classical probability the Theorem of Bayes results from the properties of the concept of conditional probability. Therefore it has to be expected that in the

[3] The Theorem of Bayes and the problem of computing posterior probabilities or posterior expectations is a frequent subject in literature dealing with generalized probability: In his fundamental book [22], Walley derived the ‘Generalized Bayes Rule’, which also is used in the robust Bayesian approach (see, f.i., [5], [24], [15], [18]).

theory of interval probability the role of conditional probability — and especially of the concept employed — proves to be decisive. The transition from prior probability to posterior probability necessarily consists of two steps:

1. Derivation of an F-probability field for which the prior is marginal probability and the conditional probability is given.
2. Derivation of conditional probability relative to the actual observation.

For the first step the method to be applied in case of interval probability is obvious: Marginal probability and conditional probability (due to the canonical concept) have to be combined by means of the Cartesian product of two structures. This generates the product rules $L(A \cap Z) = L(A) \cdot L(A \mid Z)$ and $U(A \cap Z) = U(A) \cdot U(A \mid Z)$. In Example 6 this procedure is demonstrated introducing a special case of double-dichotomy which will be employed in all of the examples to come.

Example 6 *Let the F-field describing the probability for a dichotomy of states of nature be given by* $P(Z_1) = [0.2; 0.3]$, $P(Z_2) = [0.7; 0.8]$, *and the probability of the outcome of a certain trial in case of state* Z_1 *be given by the F-field* $P(A_1 \mid Z_1) = [0.6; 0.7]$, $P(A_2 \mid Z_1) = [0.3; 0.4]$ *in case of state* Z_2 *by the F-field* $P(A_1 \mid Z_2) = [0.1; 0.2]$, $P(A_2 \mid Z_2) = [0.8; 0.9]$. *Interpreting the first of the three fields as marginal probability and the two others as conditional probability according to the canonical concept one arrives at the following components of an F-field describing the combined probability of the states and outcomes:*

$P(A_1 \cap Z_1) = [0.12; 0.21]$	$P(A_2 \cap Z_1) = [0.06; 0.12]$	$P(Z_1) = [0.2; 0.3]$
$P(A_1 \cap Z_2) = [0.07; 0.16]$	$P(A_2 \cap Z_2) = [0.56; 0.72]$	$P(Z_2) = [0.7; 0.8]$
$P(A_1)$	$P(A_2)$	$P(\Omega_4) = [1]$

This is partial determinate F-probability and the process of normal completion has to be employed in order to calculate the components $P(A_1)$ *and* $P(A_2)$. *In the present situation the results are gained easily: Let* $p(Z_1) = a$ *be a K-function belonging to the structure of the prior probability,* $p(A_1 \mid Z_1) = b$ *and* $p(A_1 \mid Z_2) = c$ *be K-functions belonging to the structures of two marginal probabilities. Therefore:* $0.2 \le a \le 0.3$; $0.6 \le b \le 0.7$; $0.1 \le c \le 0.2$. *The possible values of* $a \cdot b$ *produce* $P(A_1 \cap Z_1)$, *those of* $(1-a) \cdot c$ *produce* $P(A_1 \cap Z_2)$ *and the values of* $a \cdot b + (1-a) \cdot c$ *produce* $P(A_1)$. *It is easily controlled, that* $a = 0.2$, $b = 0.6$, $c = 0.1$ *render the minimum of* $a \cdot b + (1-a) \cdot c$, *so that* $L(A_1) = 0.20$ *results, and* $a = 0.3$, $b = 0.7$, $c = 0.2$ *render* $U(A_1) = 0.35$. *Since an F-field possesses conjugate interval limits, one arrives for* $A_1 = \neg A_2$ *at* $L(A_2) = 0.65$, $U(A_2) = 0.80$. *The last line of the table above reads:*

$$P(A_1) = [0.20; 0.35] \quad P(A_2) = [0.65; 0.80] \quad P(\Omega_4) = [1].$$

The results of the procedure described are those components of the combined F-field which are relevant with respect to posterior probability. The components still *lacking would be calculated in an analogous manner, for instance* $P[(A_1 \cap Z_1) \cup (A_2 \cap Z_2)] = [0.76; 0.86]$. □

While the canonical concept is inevitable for step 1, there is a possibility to choose between the concepts as far as step 2 is concerned: the calculation of the posterior probability for each observation. The decision in favour of the intuitive concept is quite common and promises some remarkable advantages:

1. The F-quality of the posterior probability is guaranteed.
2. The structure of this F-field can be interpreted as the Cartesian product of the structures of the marginal probability and of the conditional F-probability belonging to the actual observation.

On the other hand, use of the canonical concept includes the risk that the outcome cannot be interpreted as a generalization of a classical probability, since the resulting intervals do not define a structure.

It is therefore advisable to calculate posterior probability by means of the intuitive concept, if this posterior constitutes the only and final goal of the analysis. However, in the following it will be demonstrated, that there are good reasons for the opposite decision, if the posterior probability is to be employed as a basis for further analysis. Two situations will be considered:

1. The posterior probability of one trial is used as prior probability for another trial which is independent from the first one.
2. The posterior probability is the basis of a decision between different actions.

As to the *first of the two aspects*: In classical theory it is seen as one of the most important merits attributed to the employment of Bayes' theorem that the transition from the prior probability to the posterior is a definitive one: After the trial the posterior takes over the role of the prior. If a next trial is independent from the first one the posterior of the former trial, therefore, is the prior of the next. Obviously the following must be seen as a substantial criterion for a successful transfer of Bayes' theorem to interval probability: The results have to be the same, whether two mutually independent trials are combined to one trial, or the posterior of the first one is used as prior for the second one. It can be shown that these requirements are met, provided that the Theorem of Bayes is executed by means of the canonical concept of conditional probability. For brevity the proof will be limited to the case of two states of nature and two possible observations.

Proposition 1 *Let* (Z_1, Z_2) *be a dichotomy of the states of nature with the prior F-probability given by* $P(Z_1) = [L; U]$.

A first trial with possible outcome A_1 *or* A_2 *is characterized by F-probabilities given by* $P(A_1 \mid Z_1) = [l_{11}; u_{11}]$, $P(A_1 \mid Z_2) = [l_{21}; u_{21}]$. *A second trial which* *is independent from the first one, has the outcomes B_1 and B_2. The ruling F-probabilities are given by $P(B_1 \mid Z_1) = [l_{12}; u_{12}]$, $P(B_1 \mid Z_2) = [l_{22}; u_{22}]$. If in the Theorem of Bayes the canonical concept of conditional probability is employed, a trial which originates from a combination of the observations A. and B. renders the same posterior probability as the procedure, in which the posterior probability of the first trial is taken as prior probability for the second one.* □

For the *proof* of this proposition it is sufficient to show that both procedures produce the same probability components $P(A_i \cap B_j \cap Z_r)$ and $P(A_i \cap B_j)$, since the final probability is derived from the interval limits of these components. The demonstration will be given for $P(A_1 \cap B_1 \cap Z_r)$, $r = 1, 2$, and $P(A_1 \cap B_1)$.

1. In case of a combined trial, because of mutual independence of the trials one arrives at $P(A_1 \cap B_1 \mid Z_1) = [l_{11} \cdot l_{12}; u_{11} \cdot u_{12}]$, $P(A_1 \cap B_1 \mid Z_2) = [l_{21} \cdot l_{22}; u_{21} \cdot u_{22}]$ and together with the marginal probability of the states of nature: $P(A_1 \cap B_1 \cap Z_1) = [l_{11} \cdot l_{12} \cdot L; u_{11} \cdot u_{12} \cdot U]$, $P(A_1 \cap B_1 \cap Z_2) = [l_{21} \cdot l_{22} \cdot (1-U); u_{21} \cdot u_{22} \cdot (1-L)]$. These two components are sufficient to calculate $P(A_1 \cap B_1)$.
2. If the first trial is executed separately, conditional probability and marginal probability produce $P(A_1 \cap Z_1) = [l_{11} \cdot L; u_{11} \cdot U]$, $P(A_1 \cap Z_2) = [l_{21} \cdot (1-U); u_{21} \cdot (1-L)]$. The component of the union of these events[4] is designated by $P(A_1) = [L_1; U_1]$. The posterior probability of the first trial in case of observation A_1 — which will be used as prior for the second trial — is defined by the canonical conditional probability as $P(Z_1 \mid A_1) = \left[\frac{l_{11} \cdot L}{L_1}; \frac{u_{11} \cdot U}{U_1}\right]$, $P(Z_2 \mid A_1) = \left[\frac{l_{21} \cdot (1-U)}{L_1}; \frac{u_{21} \cdot (1-L)}{U_1}\right]$. Hence, conditional to A_1 the probability-components for the observation B_1 of the second trial read as $P(B_1 \cap Z_1 \mid A_1) = \left[\frac{l_{12} \cdot l_{11} \cdot L}{L_1}; \frac{u_{12} \cdot u_{11} \cdot U}{U_1}\right]$, $P(B_1 \cap Z_2 \mid A_1) = \left[\frac{l_{22} \cdot l_{21} \cdot (1-U)}{L_1}; \frac{u_{22} \cdot u_{21} \cdot (1-L)}{U_1}\right]$. In order to arrive at the components of the events $A_1 \cap B_1 \cap Z_1$ and $A_1 \cap B_1 \cap Z_2$, canonical conditional and marginal probability must be combined:
   $L(A_1 \cap B_1 \cap Z_1) = L(B_1 \cap Z_1 \mid A_1) \cdot L(A_1) = \frac{l_{11} \cdot l_{12} \cdot L}{L_1} \cdot L_1 = l_{11} \cdot l_{12} \cdot L$,
   $U(A_1 \cap B_1 \cap Z_1) = U(B_1 \cap Z_1 \mid A_1) \cdot U(A_1) = \frac{u_{11} \cdot u_{12} \cdot U}{U_1} \cdot U_1 = u_{11} \cdot u_{12} \cdot U$
   and corresponding procedures for Z_2. Both components are equal to those resulting from the combined trial and consequently as well $P(A_1 \cap B_1)$ as the canonical conditional probability are alike: Both methods produce the same posterior. □

In Example 7 this equivalence is demonstrated in the case of the F-probability field introduced in Example 6.

Example 7 *For the prior probability and the conditional probability of Example 6 the posterior probability — defined by the canonical concept results as*

$$P(Z_1 \mid A_1) = \left[\frac{0.12}{0.20}; \frac{0.21}{0.35}\right] = [0.60; 0.60]$$
$$P(Z_2 \mid A_1) = \left[\frac{0.07}{0.20}; \frac{0.16}{0.35}\right] = [0.35; 0.46].$$

These two components can be interpreted as R-probability, since $p(Z_1|A_1) = 0.60$, $p(Z_2 \mid A_1) = 0.40$, is a K-function in accordance with all

[4]The appropriate method of calculation is demonstrated in Example 6.

interval limits. It will be seen that despite the lack of F-quality this assignment can be used as a prior for a next trial. Let

$$P(B_1 \mid Z_1) = [0.7; 0.9] \qquad P(B_2 \mid Z_1) = [0.1; 0.3]$$
$$P(B_1 \mid Z_2) = [0.2; 0.4] \qquad P(B_2 \mid Z_2) = [0.6; 0.8].$$

Combined with the new prior produced by observation A_1:

$$\begin{array}{ll|l} P(B_1 \cap Z_1 \mid A_1) = [0.42; 0.54] & P(B_2 \cap Z_1 \mid A_1) = [0.06; 0.18] & P(Z_1 \mid A_1) = [0.60; 0.60] \\ P(B_1 \cap Z_2 \mid A_1) = [0.07; 0.184] & P(B_2 \cap Z_2 \mid A_1) = [0.21; 0.37] & P(Z_2 \mid A_1) = [0.35; 0.46] \end{array}$$

In order to calculate components of the absolute probability, the component $P(A_1) = [0.20; 0.35]$ according to Example 6 has to be multiplied — which is executed only for the events produced by observation B_1:

$$L(A_1 \cap B_1 \cap Z_1) = 0.42 \cdot 0.20 = 0.084 \qquad U(A_1 \cap B_1 \cap Z_1) = 0.54 \cdot 0.35 = 0.189$$
$$L(A_1 \cap B_1 \cap Z_2) = 0.07 \cdot 0.20 = 0.014 \qquad U(A_1 \cap B_1 \cap Z_2) = 0.184 \cdot 0.35 = 0.064.$$

If, on the other hand, the mutually independence trials were combined, the components of $A_1 \cap B_1$ would be:

$$\begin{aligned} P(A_1 \cap B_1 \mid Z_1) &= [0.6 \cdot 0.7; 0.7 \cdot 0.9] = [0.42; 0.63] \\ P(A_1 \cap B_1 \mid Z_2) &= [0.1 \cdot 0.2; 0.2 \cdot 0.4] = [0.02; 0.08]. \end{aligned}$$

With respect to the marginal probability $P(Z_1) = [0.2; 0.3]$, $P(Z_2) = [0.7; 0.8]$ the outcome of the combined trial is partially described by the components

$$\begin{aligned} P(A_1 \cap B_1 \cap Z_1) &= [0.42 \cdot 0.2; 0.63 \cdot 0.3] = [0.084; 0.189] \\ P(A_1 \cap B_1 \cap Z_2) &= [0.02 \cdot 0.7; 0.08 \cdot 0.8] = [0.014; 0.064] \end{aligned}$$

demonstrating the conformity of the two procedures with regard to probability of the observations. □

It should be noted that the procedure described in Proposition 1 and demonstrated in Example 7 is not a mere transfer of the procedures customary in classical theory. The posterior probability resulting from the first trial is conditional probability relative to the actual observation. Prior probability is always marginal probability, hence total probability, not a conditional one. Therefore total probability has to be reconstructed by means of the marginal component of the actual observation in the first trial. This step does not influence the result in classical theory — and is left out therefore — but it is inevitable for interval probability!

Concerning the *decision-theoretic approach* it has been shown recently ([1]) that with regard to the optimization of decisions in the general case of interval probability the Theorem of Bayes — at least as far as it employs the intuitive concept — does not render what its counterpart for classical probability renders: that the Bernoulli-optimal action with respect to the posterior probability generated by the actual observation produces the corresponding branch of the optimal decision function. Hence the so called 'Main Theorem of Bayesian Decision Analysis' does not hold for interval probability.

It can be demonstrated that in the general case of interval probability this phenomenon is inevitable — beyond all questions about the methodology of Bayes' theorem. In classical theory the branch of a decision function attributed to a certain observation produces an expected gain not depending on the circumstances related to the other possible observations. Therefore this expectation can be compared directly with those of respective branches belonging to other — competing — decision functions, a task, which is achieved easily via the Theorem of Bayes.

In presence of ambiguity the situation is different: If the expected gain of a decision function is calculated, each of the partial sums generated by an observation can be influenced by circumstances which originally refer to any of the other possible observations.This is a rule of thumb for decision functions:

Classical probability — only the actual observation counts.

Interval probability — all possible observations count.

Example 8, related to Examples 6 and 7, shows: If two gain functions differ only for observation A_2, nevertheless the contribution of observation A_1 to the interval expectation of the total gain may be influenced by this difference.

Example 8 *Z_1, Z_2 are two states of nature and A_1, A_2 are two possible observations, where the marginal probability $P(Z_1)$, $P(Z_2)$ and the canonical conditional probabilities $P(A_1 \mid Z_1)$, $P(A_2 \mid Z_1)$, $P(A_1 \mid Z_2)$, $P(A_2 \mid Z_2)$ are given in Example 1. Remember, that for K-functions of the respective structures*

$$p(Z_1) = a, \quad p(A_1 \mid Z_1) = b, \quad p(A_1 \mid Z_2) = c$$

the interval limits are given by

$$0.2 \leq a \leq 0.3, \quad 0.6 \leq b \leq 0.7, \quad 0.1 \leq c \leq 0.2.$$

The structure of the resulting F-field then consists of K-functions with the components given by

$p(A_1 \cap Z_1) = a \cdot b$	$p(A_2 \cap Z_1) = a \cdot (1-b)$	$p(Z_1) = a$
$p(A_1 \cap Z_2) = (1-a) \cdot c$	$p(A_2 \cap Z_2) = (1-a) \cdot (1-c)$	$p(Z_2) = 1-a$
$p(A_1) = a \cdot b + (1-a) \cdot c$	$p(A_2) = a \cdot (1-b) + (1-a) \cdot (1-c)$	$p(\Omega_4) = [1]$

producing the interval limits of these components as

$P(A_1 \cap Z_1) = [0.12; 0.21]$	$P(A_2 \cap Z_1) = [0.06; 0.12]$	$P(Z_1) = [0.2; 0.3]$
$P(A_1 \cap Z_2) = [0.07; 0.16]$	$P(A_2 \cap Z_2) = [0.56; 0.72]$	$P(Z_2) = [0.7; 0.8]$
$P(A_1) = [0.20; 0.35]$	$P(A_2) = [0.65; 0.80]$	$P(\Omega_4) = [1]$

A first decision function $D_1(.)$ is characterized by the following gains

$$D_1(A_1 \cap Z_1) = 4 \qquad D_1(A_2 \cap Z_1) = 6$$
$$D_1(A_1 \cap Z_2) = 8 \qquad D_1(A_2 \cap Z_2) = 2.$$

The expected gain $e(D_1(.))$ for a K-function described by a, b and c is given as

$$e(D_1(.)) = 4ab + 8(1-a)c + 6a(1-b) + 2(1-a)(1-c) = 2 + 2a(2-b-3c) + 6c.$$

Since $e(D_1(.))$ is minimal for $a = 0.2$, $b = 0.7$, $c = 0.1$ and $\mathbb{E}(D_1(.)) = [3.0; 3.68]$. For every K-function, $e(D_1(.))$ can be divided into the two branches: $e(D_1(.)) = e(D_1(.) \cap A_1) + e(D_1(.) \cap A_2)$ where

$$e(D_1(.) \cap A_1) = 4ab + 8(1-a)c,\ e(D_1(.) \cap A_2) = 6a(1-b) + 2(1-a)(1-c).$$

With respect to the roles of the two branches in determining $e(D_1(.))$ they have to be evaluated in the same way as $e(D_1(.))$ itself, i.e., using $a = 0.2$, $b = 0.7$, $c = 0.1$ to produce the two parts of $\mathbb{L}(D_1(.)) = 3.0$: $\mathbb{L}^*(D_1(.) \cap A_1) = 1.20$, $\mathbb{L}^*(D_1(.) \cap A_2) = 1.80$, and $a = 0.3$, $b = 0.6$, $c = 0.2$ to produce the respective parts of $\mathbb{U}(D_1(.)) = 3.68$: $\mathbb{U}^*(D_1(.) \cap A_1) = 1.84$, $\mathbb{U}^*(D_1(.) \cap A_2) = 1.84$. As far as comparisons with other decision functions are concerned, the branch of $D_1(.)$ determined by the observation A_1 therefore is represented by $[1.20; 1.84]$. Now let a second decision function $D_2(.)$ be given by

$$\begin{array}{ll} D_2(A_1 \cap Z_1) = 4 & D_2(A_2 \cap Z_1) = 3 \\ D_2(A_1 \cap Z_2) = 8 & D_2(A_2 \cap Z_2) = 2. \end{array}$$

This leads to

$$e(D_2(.)) = 4ab + 8(1-a)c + 3a(1-b) + 2(1-a)(1-c) = 2 + a(1+b-6c) + 6c$$

and this is minimal for $a = 0.2$, $b = 0.6$, $c = 0.1$ and maximal for $a = 0.3$, $b = 0.7$, $c = 0.2$, producing $\mathbb{E}(D_2(.)) = [2.8; 3.35]$. If this interval expectation is divided into the two branches generated by the observation of A_1 and A_2, one arrives at

$$e(D_2(.) \cap A_1) = 4ab + 8(1-a)c, \quad e(D_2(.) \cap A_2) = 3a(1-b) + 2(1-a)(1-c)$$

together with the results for $a = 0.2$, $b = 0.6$, $c = 0.1$: $\mathbb{L}^*(D_2(.) \cap A_1) = 1.12$, $\mathbb{L}^*(D_2(.) \cap A_2) = 1.68$, and for $a = 0.3$, $b = 0.7$, $c = 0.2$: $\mathbb{U}^*(D_2(.) \cap A_1) = 1.96$, $\mathbb{U}^*(D_2(.) \cap A_2) = 1.39$.

There are two striking findings:

1. *$\mathbb{U}^*(D_2(.) \cap A_2) < \mathbb{L}^*(D_2(.) \cap A_2)$. Obviously $\mathbb{L}^*(D_2(.) \cap A_2)$ and $\mathbb{U}^*(D_2(.) \cap A_2)$ may not be confounded with the lower and upper interval limits for the expectation of $D_2(.) \cap A_2$, which can be calculated as $\mathbb{L}(D_2(.) \cap A_2) = 1.39$ (produced by $a = 0.3$, $b = 0.7$, $c = 0.2$) and $\mathbb{U}(D_2(.) \cap A_2) = 1.68$ (produced by $a = 0.2$, $b = 0.6$, $c = 0.1$). In the case of decision function $D_2(.)$ therefore that constellation of K-functions, which leads to the maximal $e(D_2(.))$, results in the smallest possible value of $e(D_2(.) \cap A_2)$, and that constellation, which minimizes $e(D_2(.))$, happens to maximize the value of $e(D_2(.) \cap A_2)$.*

2. $\mathbb{L}^*(D_2(.) \cap A_1) \neq \mathbb{L}^*(D_1(.) \cap A_1)$ *and* $\mathbb{U}^*(D_2(.) \cap A_1) \neq \mathbb{U}^*(D_1(.) \cap A_1)$. *Both interval limits describing the contribution of branch* A_1 *to the expected gain are different for decision function* $D_1(.)$ *and decision function* $D_2(.)$ *— although all of the data describing branch* A_1 *are equal for both decision functions. The differences between the contributions of branch* A_1 *are caused by differences concerning the gains in case of observation* A_2. □

This phenomenon demonstrates the impossibility of qualifying the contribution of the branch attributed to the actual observation only by the circumstances of this observation without consideration of data related to other possible observations. In interval probability a decision function can only be judged or compared with others as a whole — not piecewise for each branch separately. Any kind of Theorem of Bayes, however, bases its calculation of the posterior probability only upon the circumstances of the actual observation — irrespective of the circumstances relating to other observations. Therefore no *posterior probability contains enough information* to qualify a branch of a decision function in comparison with the corresponding branches of competing decision functions.

The situation is different, if the problem considered is characterized by a very special type of gain function: Gains different from zero are supposed to be possible only if the actual observation is A_1. Therefore decision functions $D(. \cap .)$ are admissible for competition only if satisfying the requirements $D(A_i \cap Z_j) = 0, \ \forall i \neq 1, \forall j$. In this case the expected total gain and the expected gain for the branch A_1 are identical for every K-function: $e(D(.)) = e(D(.) \cap A_1)$. Consequently the following relations hold: $\mathbb{L}(D(.) \cap A_1) = \mathbb{L}(D(.))$ and $\mathbb{U}(D(.) \cap A_1) = \mathbb{U}(D(.))$. While at first this assumption seems to be very unrealistic, its systematic application to every actual observation A_i — instead of A_1 — generates a strategy which obviously is suboptimal in the general case, but may be understood as a kind of approximation to the optimal strategy: For each actual observation A_i that action $D(.) \cap A_i$ is chosen, which is best w.r.t. $[\mathbb{L}(D(.) \cap A_i); \mathbb{U}(D(.) \cap A_i)]$, irrespective of all observations which could have been made and the gains which would have been possible, if this observations had occured.

This strategy is much simpler than that founded on the complete decision function. It is an imitation of the proceeding in classical probability. In Example 9 it is demonstrated using the data of Example 8.

Example 9 *In the case of observation* A_1 *for the branch* $D_1(A_1 \cap Z_1) = D_2(A_1 \cap Z_1) = 4$, $D_1(A_1 \cap Z_1) = D_2(A_1 \cap Z_2) = 8$ *the decisive interval-expectation is given by*

$$\mathbb{L}(D_1(.) \cap A_1) = \mathbb{L}(D_2(.) \cap A_1) = 1.12 \quad (a = 0.2, b = 0.6, c = 0.1)$$
$$\mathbb{U}(D_1(.) \cap A_1) = \mathbb{U}(D_2(.) \cap A_1) = 1.96 \quad (a = 0.3, b = 0.7, c = 0.2).$$

In case of observation A_2*: For* $D_1(A_2 \cap Z_1) = 6$, $D_1(A_2 \cap Z_2) = 2$ *one arrives at*

$$\mathbb{L}(D_1(.) \cap A_2) = 1.64 \quad (a = 0.2, b = 0.7, c = 0.2)$$
$$\mathbb{U}(D_1(.) \cap A_2) = 1.98 \quad (a = 0.3, b = 0.6, c = 0.1),$$

for $D_2(A_2 \cap Z_1) = 3$, $D_2(A_2 \cap Z_2) = 2$*:*

$$\mathbb{L}(D_2(.) \cap A_2) = 1.39 \quad (a = 0.3, b = 0.7, c = 0.2)$$
$$\mathbb{U}(D_2(.) \cap A_2) = 1.68 \quad (a = 0.2, b = 0.6, c = 0.1). \qquad \square$$

Two remarks are useful:

Expectations belonging to different observations are based on contradictory assumptions. Therefore they are not suitable for being combined.

Comparison of actions which are characterized by means of interval expectation depends on the attitude of the decision-maker towards ambiguity. It may be described by the choice of $\eta L(G) + (1-\eta)U(G)$, $0 \leq \eta \leq 1$, as the decisive quantity. Since it can be understood, that the larger value of gain G always is preferred, η is interpreted as a measure of caution.

Because of the goal of this section it is asked whether a posterior probability generated by the Theorem of Bayes can be employed in calculating the expectation $[\mathbb{L}(D(.) \cap A_1); \mathbb{U}(D(.) \cap A_1)]$ produced by the actual observation A_1.

Using again the data of Example 9 it will be demonstrated in Example 10 that with respect to that type of Theorem of Bayes, which employs the intuitive concept of conditional probability, the answer to this question must be negative.

Example 10 *The intuitive conditional probability* $iP(Z_1 \mid A_1)$*,* $iP(Z_2 \mid A_1)$ *obviously is determined by* $iL(Z_1 \mid A_1) = \min\limits_{\mathcal{M}} \frac{ab}{ab+(1-a)c}$*,* $iU(Z_1 \mid A_1) = \max\limits_{\mathcal{M}} \frac{ab}{ab+(1-a)c}$ *with*

$$\mathcal{M} = \{p_{a,b,c}(.); 0.2 \leq a \leq 0.3; 0.6 \leq b \leq 0.7; 0.1 \leq c \leq 0.2\}.$$

It is easily seen, that the minimum is produced by $a = 0.2$*,* $b = 0.6$*,* $c = 0.2$ *and the maximum by* $a = 0.3$*,* $b = 0.7$*,* $c = 0.1$*. The resulting i-conditional F-probability field is given by* $iP(Z_1 \mid A_1) = [0.429; 0.750]$*,* $iP(Z_2 \mid A_1) = [0.250; 0.571]$*. The i-conditional expectation of the gain function produced by the decision function* $D(.) = D_1(.) = D_2(.)$ *with* $D(A_1 \cap Z_1) = 4$*,* $D(A_1 \cap Z_2) = 8$ *is determined by*

$$i\mathbb{L}(D(.) \mid A_1) = 0.750 \cdot 4 + 0.250 \cdot 8 = 5$$
$$i\mathbb{U}(D(.) \mid A_1) = 0.429 \cdot 4 + 0.571 \cdot 8 = 7.429.$$

To achieve the interval-expectation of $D(.) \cap A_1$*, conditional expectation must be combined with the corresponding component of marginal probability:* $P(A_1) = [0.20; 0.35]$*. Therefore* $i\mathbb{L}(D(.) \cap A_1) = 5 \cdot 0.20$*,* $i\mathbb{U}(D(.) \cap A_1) = 7.429 \cdot 0.35$ *and* $i\mathbb{E}(D(.) \cap A_1) = [1.00; 2.60]$ *instead of the true interval expectation, as calculated in Example 9:* $\mathbb{E}(D(.) \cap A_1) = [1.12; 1.96]$*. Like in other situations, employment of the intuitive concept generates a loss in sharpness of the result.* $\square$

If, however, the canonical concept is applied, the conditional expectation of the gain produced by the decision function $D(.)$ for the observation A_1, due to the definition described in Section 2, reads as $\mathbb{E}(D(.) \mid A_1) = \left[\frac{\mathbb{L}(D(.) \cap A_1)}{L(A_1)}; \frac{\mathbb{U}(D(.) \cap A_1)}{U(A_1)}\right]$.

Combined with the component $[L(A_1); U(A_1)]$ of the marginal probability this conditional expectation produces $\mathbb{E}(D(.) \cap A_1) = [\mathbb{L}(D(.) \cap A_1); \mathbb{U}(D(.) \cap A_1)]$, due to the simplified optimal strategy, as it was described above (Example 11).

It is, therefore, justified to use the designation 'Interval Bayes-Strategy' for the method of selecting in case of observation A_1 that action which produces the largest expected gain — judged by means of the individual caution — with respect to the posterior probability generated by the Theorem of Bayes with the canonical concept of conditional probability.

Example 11 *Because of* $\mathbb{E}(D(.) \cap A_1) = [1.12; 1.96]$ *(Example 9) and* $P(A_1) = [0.20; 0.35]$ *(Example 8), the conditional expectation results as* $\mathbb{L}(D(.) \mid A_1) = \frac{1.12}{0.20}$, $\mathbb{U}(D(.) \mid A_1) = \frac{1.96}{0.35}$ *or* $\mathbb{E}(D(.) \mid A_1) = [5.60; 5.60]$. *The quality of the result is not affected by the fact, that this interval possesses length zero.* □

Hence, use of the canonical concept allows the Interval Bayes-Strategy, distinguished from the strategy based upon the optimal decision function only by neglecting any information concerning observations which did not occur. If the omission of such 'counterfactual information' is accepted on principle, the Interval Bayes-Strategy must be regarded as optimal.

7 Conclusions

This paper contributes to the question of defining conditional interval probability appropriately. A symbiosis of the intuitive and the canonical concept of conditional probability is proposed, resulting in recommendations which of the concepts should be used for what propose.

The results of Sections 3–6 can for short be interpreted to favour the employment of the intuitive concept in any situation where conditional probability is seen as a goal in itself, therefore in updating, whether it is achieved directly or by means of the Theorem of Bayes: the final result should be described by the intuitive concept of conditional probability.

The canonical concept proves to be superior always when conditional probability is used as a tool for further analysis. This applies to the transfer of information from one model to another and to the derivation of a posterior probability by means of the Theorem of Bayes, if this posterior is employed as prior for an independent trial, or as basis for decisions between possible actions. Additionally this concept produces a Theorem of Total Probability and the consistency with marginal probability in the case of independence as defined by strong extension.

While the intuitive concept guarantees that its outcome describing the final result of an analysis always can be interpreted as interval probability or as interval expectation, it is possible that the outcome of the canonical concept, which is employed as a tool for further calculations, does not possess the F- or even R-quality, resp., the quality of interval expectation, without loss of usefulness: an obvious analogy to the role of complex numbers in algebra.

References

[1] T. Augustin. On the suboptimality of the Generalized Bayes Rule and robust Bayesian procedures from the decision theoretic point of view — a cautionary note on updating imprecise priors. `www.stat.uni-muenchen.de/~thomas/robust-bayes-suboptimal.pdf`, 2003.

[2] T. Augustin and F. Coolen. Nonparametric predictive inference and interval probability. To appear in: *Journal of Statistical Planning and Inference*, 2003.

[3] J.M. Bernard. Non-parametric inference about an unknown mean using the imprecise Dirichlet model. *In:* [9], 40–50.

[4] I. Couso, S. Moral, and P. Walley. A survey of concepts for imprecise probabilities. *Risk Decision and Policy*, 5: 165–181, 2000.

[5] F.G. Cozman. Computing posterior upper expectations. *In:* G. de Cooman, F.G. Cozman, S. Moral, and P. Walley (eds.), *ISIPTA '99: Proceedings of the First International Symposium on Imprecise Probabilities and their Applications*. University of Ghent, 131–140, 1999.

[6] F.G. Cozman. Constructing sets of probability measures through Kuznetsov's independence condition. *In:* [9], 104–111.

[7] F. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *In:* [9], 112–121.

[8] G. de Cooman. Integration and conditioning in numerical possibility theory. *Annals of Mathematics and Artificial Intelligence*, 32: 87–123, 2001.

[9] G. de Cooman, T. Fine, S. Moral, and T. Seidenfeld (eds.). *ISIPTA01: Proceedings of the Second International Symposium on Imprecise Probabilities and their Applications*. Cornell University, Ithaca (N.Y.). Shaker, Maastricht, 2001.

[10] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38: 325–339, 1967.

[11] D. Dubois and H. Prade. Focusing versus updating in belief function theory. *In:* R.R. Yager, M. Fedrizzi, and J. Kacprzyk (eds.), *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, New York, 71–95, 1994.

[12] J.Y. Halpern and R. Fagin. Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54: 275–317, 1992.

[13] T. Fetz. *Sets of joint probability measures generated by random sets*. Doctoral Thesis, University of Innsbruck, 2002.

[14] V.P. Kuznetsov. *Interval Statistical Methods* (in Russian). Radio i Svyaz Publ., Moscow, 1991.

[15] M. Lavine. Sensitivity in Bayesian statistics, the prior and the likelihood. *Journal of the American Statistical Association*, 86 (414): 396–399, 1991.

[16] E. Miranda and G. de Cooman. Independent products of numerical possibility measures. *In:* [9], 237–246.

[17] S. Moral. Epistemic irrelevance on sets of desirable gambles. *In:* [9], 247–254.

[18] D. Rios Insua and F. Ruggeri (eds.). *Robust Bayesian Analysis*. Lecture Notes in Statistics 152, Springer, New York, 2000.

[19] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[20] B. Vantaggi. Graphical models for conditional independence structures. *In:* [9], 332–341.

[21] P. Vicig. Epistemic independence for imprecise probabilities. *International Journal of Approximate Reasoning*, 24: 235–250, 2000.

[22] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, New York, 1991.

[23] P. Walley. Inferences from multinomial data: Learning from a bag of marbles (with discussion). *Journal of the Royal Statistical Society*, B 58: 3–57, 1996.

[24] L. Wasserman. Bayesian robustness. *In:* S. Kotz, C.B. Read, and D.L. Banks (eds.), *Encyclopedia of Statistical Sciences, Update Volume 1*. Wiley, New York, 45–51, 1997.

[25] K. Weichselberger. Axiomatic foundations of the theory of interval-probability. *In:* V. Mammitzsch and H. Schneeweiß (eds.), *Symposia Gaussiana, Proceedings of the 2nd Gauss-Symposium, Conference B*. De Gruyter, Berlin, 47–64, 1995.

[26] K. Weichselberger. The theory of interval probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24: 149–170, 2000.

[27] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I — Intervallwahrscheinlichkeit als umfassendes Konzept*, in cooperation with T. Augustin and A. Wallner. Physica, Heidelberg, 2001.

[28] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung II — Die Theorie von Intervallwahrscheinlichkeit*. In preparation.

[29] N. Wilson. Modified upper and lower probabilities based on imprecise likelihoods. *In:* [9],

[30] C. Yu and F. Arasta. On conditional belief functions. *International Journal of Approximate Reasoning*, 10: 155–172, 1994.

[31] M. Zaffalon. Statistical inference of the naive credal classifier. *In:* [9], 384–393.

K. Weichselberger and T. Augustin are with the Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstr. 33, 80539 München, Germany. E-mail: {weichsel, augustin}@stat.uni-muenchen.de