Augustin, Coolen:

Nonparametric predictive inference and interval probability

Sonderforschungsbereich 386, Paper 257 (2001)

Nonparametric predictive inference and interval probability*

T. Augustin
Ludwig-Maximilians University
Department of Statistics
Akademiestr. 1
D-80799 Munich, Germany
thomas@stat.uni-muenchen.de

F.P.A. Coolen
University of Durham
Department of Mathematical Sciences
Science Laboratories, South Road
Durham, DH1 3LE, England
Frank.Coolen@durham.ac.uk

Abstract

This paper presents the unique position of $A_{(n)}$-based nonparametric predictive inference within the theory of interval probability. It provides a completely new understanding, leading to powerful new results and a well-founded justification of such inferences by proving strong internal consistency results.

Keywords: $A_{(n)}$; Capacities; Conditioning; Consistency; Imprecise probabilities; Interval probability; Nonparametrics; Predictive inference; Updating.

MSC: 62A99 – Foundations; 62G99 – Nonparametric inference

*We gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft in the frame of the Sonderforschungsbereich 386 and from the London Mathematical Society for research visits to Durham and Munich, which were of great benefit to this collaboration.

1 Introduction

The assumption $A_{(n)}$ was proposed by Hill [25, 26] for prediction in the case of extremely vague a priori knowledge about the form of the underlying distribution. Let x_i, $i = 1, \ldots, n$, be the data values obtained in sampling from a finite population, and let $x_{(i)}$ be their ordered values in increasing order of magnitude. Let X_i be the corresponding pre-data random quantities, so that the data consist of the realized values, $X_i = x_i$, $i = 1, \ldots, n$. Following Hill [27], we define $A_{(n)}$ as follows.

1. The observable random quantities $X_1, \ldots, X_n$ are exchangeable. (In the original definition of $A_{(n)}$ [25], exchangeability was not included allowing more general situations.)

2. Ties have probability 0. (Generalization to include possible ties is straightforward, see Hill [26], but leads to more awkward notation.)

3. Given data x_i, $i = 1, \ldots, n$, the probability that the next observation falls in the open interval $I_j = (x_{(j-1)}, x_{(j)})$ is $1/(n+1)$, for each $j = 1, \ldots, n+1$, where we define $x_{(0)} = -\infty$ and $x_{(n+1)} = \infty$.

It is clear that $A_{(n)}$ is a post-data assumption related to finite exchangeability [21], see Hill [26] for a detailed presentation and discussion of $A_{(n)}$, and an overview of related work, including important contributions by Dempster [15] and Lane and Sudderth [30].

The assumption $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest in inference based on such data. However, it does provide bounds for probabilities, by what is essentially an application of De Finetti's 'fundamental theorem of probability' [21]. Theory of interval probability makes it clear that such bounds contain valuable information, both on uncertainty of events and on indeterminacy caused by restricted information. Therefore, we study the position of these bounds in the theory of interval probability [36, 40]. In this paper, we prove that $A_{(n)}$-based interval probabilities have a position that is so-far unique in the theory of interval probability.

De Finetti's [21] representation theorem uses a similar setting to justify a Bayesian framework to learn about an underlying parameter, and a probability distribution for that parameter, but he relies on the assumption that indeed there is an infinite sequence of random quantities involved, whereas our interest here is mostly in inference on a single future observation. Even more, the Bayesian approach as justified by De Finetti's [21] important results, explicitly needs a specified prior distribution, and together with the conditional independence of future observations (conditional on an unknown parameter) this adds quite a bit more structure to the data then our inference based on $A_{(n)}$. Such inferences have a predictive and nonparametric nature, and seem suitable if there is hardly any knowledge about the random quantities of interest, other than the first n observations, or, which may be more realistic, if one explicitly does not want to use such information. This may occur, for example, if one wants to study the (often hidden) effect of additional structural assumptions underlying statistical models or methods. Inferences based on such restricted knowledge have also been called 'low structure inferences' [23] and 'black-box inferences' [30].

Recent suggestions of applications of $A_{(n)}$-based nonparametric predictive inference, for example in process control [2], condition monitoring [12] and selection [13], have also inspired this study of the theoretical properties of the interval probabilities occurring in this approach. The results in this paper support this inferential method further by showing its internal consistency from a static as well as a dynamic point of view.

The outline of this paper is as follows. Section 2 provides a brief overview of some basic aspects of interval probability, mostly based on the theory of Weichselberger [39, 40]. Section 3 presents the general form of the predictive interval probabilities as derived from the assumption $A_{(n)}$, together with some basic results that make clear their relation to other generalizations of probability theory that have been suggested in the literature. Sections 4 and 5 discuss conditioning and updating, respectively, and strong internal consistency results are derived. Finally, section 6 briefly discusses some further related aspects, including some topics for future research.

2 Some basic aspects of interval probability

The idea to use interval-valued probabilities dates back at least to the middle of the nineteenth century [6]. Since then, interval probabilities, also known as imprecise probabilities, have been suggested in quite different areas of statistics. For example, they arise naturally in several approaches to predictive inference such as Dempster's multivalued mappings [16] and Hampel's successful bets [24], in modelling uncertain knowledge in artificial intelligence [41], in economic decision theory [9], and in robust Neyman-Pearson-testing [4, 29]. Furthermore, there is a strong connection to robust Bayesian inference [5, 32]. Recently, there has been increasing activity in this area by researchers from widely varying backgrounds, resulting in a series of conferences [18, 19], special issues of journals [7, 14, 17] and a project webpage [20].

Fine and collaborators [31, 37] established a frequentist theory of interval probability. Extending De Finetti's theory [21] to interval-valued previsions, Walley [36] provides a rigorous generalization of the concept of probability based on a behavioural interpretation of subjective interval probability as bets with eventually differing maximum buying price and minimum selling price. A formal foundation of interval probability in the spirit of Kolmogorov's axioms, and not dependent on interpretation, is developed by Weichselberger [40] (see [38, 39] for some selected aspects). We use Weichselberger's concept throughout this paper, and we comment briefly on the relation to work by others, in particular Walley's theory and Choquet capacities.

According to Weichselberger [40], an axiomization of interval probability can be achieved by supplementing Kolmogorov's axioms.

Definition 1

Let $(\Omega, \mathcal{A})$ be a measurable space.

- *A set-function $p(\cdot)$ on $\mathcal{A}$ satisfying Kolmogorov's axioms is called* classical probability*; the set of all classical probabilities on $(\Omega, \mathcal{A})$ will be denoted by $\mathcal{K}(\Omega, \mathcal{A})$.*
- *A function $P(\cdot)$ on $\mathcal{A}$ is called* F-probability *with* structure $\mathcal{M}$*, if*
  - * *$P(\cdot)$ is of the form*

$$\begin{array}{rcccl} P & : & \mathcal{A} & \to & \{[L;U] \mid 0 \leq L \leq U \leq 1\} \\ & & A & \mapsto & P(A) = [L(A); U(A)] \,, \end{array} \tag{1}$$

    and

  - * *the set*

$$\mathcal{M} := \{p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid L(A) \leq p(A) \leq U(A), \quad \forall A \in \mathcal{A}\} \neq \emptyset, \tag{2}$$

    and

  - *

$$\left.\begin{array}{l} \inf\limits_{p(\cdot)\in\mathcal{M}} p(A) = L(A) \\[2ex] \sup\limits_{p(\cdot)\in\mathcal{M}} p(A) = U(A) \end{array}\right\} \quad \forall A \in \mathcal{A}\,. \tag{3}$$

Throughout this paper, the notation $P(\cdot)$ is used for interval assignments, and $p(\cdot)$ for classical probability. The property (3) is considered by several authors, for instance [29]. On finite $\sigma-$fields, it coincides with Fine's notion of envelopes [31, 37] and with Walley's concept of coherence [36].

Some consequences of Definition 1 are that for every F-probability, $L(\cdot)$ and $U(\cdot)$ are *conjugated*,

$$L(A) = 1 - U(\neg A), \quad \forall A \in \mathcal{A}\,, \tag{4}$$

and $L(\cdot)$ is superadditive and $U(\cdot)$ is subadditive,

$$L(A{\cup}B) \geq L(A){+}L(B) \quad \text{and} \quad U(A{\cup}B) \leq U(A){+}U(B), \qquad \forall A, B \in \mathcal{A}, A{\cap}B = \emptyset. \tag{5}$$

Conjugacy ensures that every F-probability is uniquely characterized by $L(\cdot)$. We call $\mathcal{F} = (\Omega; \mathcal{A}; L(\cdot))$ an *F-probability field.*

Not every set of classical probabilities is a structure, but every set can be used to construct the smallest structure containing it and a unique corresponding F-probability field. The proof of the following lemma is directly obtained from Definition 1.

Lemma 1
Consider a measurable space $(\Omega, \mathcal{A})$ and a non-empty set $\mathcal{M}_0$ of classical probabilities on $(\Omega, \mathcal{A})$. Then $P(\cdot) := [L_0(\cdot); U_0(\cdot)]$ with

$$L_0(A) := \inf_{p(\cdot)\in\mathcal{M}_0} p(A) \quad \textit{and} \quad U_0(A) := \sup_{p(\cdot)\in\mathcal{M}_0} p(A) \tag{6}$$

is F-probability*. For its structure $\mathcal{M}$ the relation $\mathcal{M} \supset \mathcal{M}_0$ holds.*

Property (3), which is essential for F-probability, is not new in the literature, but usually it is derived as a consequence of two- or totally-monotone capacities. Assignments satisfying properties beyond F-probability, in particular monotonicity properties of $L(\cdot)$, become special cases with some mathematical particularities, but without any principal superiority from the foundational point of view [3, 35, 40].

Definition 2
Let $(\Omega, \mathcal{A})$ be a measurable space. F-probability

$$\begin{array}{rcrcl} P & : & \mathcal{A} & \to & \{[L;U] \mid 0 \leq L \leq U \leq 1\} \\ & & A & \mapsto & [L(A); U(A)] \end{array}$$

is called

1. C-probability*, if $L(\cdot)$ is* two-monotone*, i.e. if*

$$L(A \cup B) + L(A \cap B) \geq L(A) + L(B), \qquad \forall A, B \in \mathcal{A}, \tag{7}$$

2. totally-monotone C-probability*, if $L(\cdot)$ is* totally-monotone*, i.e. if for all $q \in I\!N$ and for all $A_1, \ldots, A_q \in \mathcal{A}$,*

$$L\left(\bigcup_{i=1}^{q} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,q\}} (-1)^{|I|+1}\, L\left(\bigcap_{i\in I} A_i\right). \tag{8}$$

Then the F-probability field $\mathcal{C} = (\Omega; \mathcal{A}; L(\cdot))$ is called (totally-monotone) C-probability field.

Obviously, every totally-monotone C-probability is C-probability. The notion of (totally-monotone) C-probability is motivated by Choquet's [10] system for classifying set-functions.

Definition 3

Let $(\Omega, \mathcal{O})$ be a topological space and $\mathcal{A}$ its Borel σ-field. A set-function $L(\cdot)$ is called a two- *or* totally-monotone capacity, *if it satisfies (7) or (8), respectively, together with the conditions*

$$L(\emptyset) = 0\,, \quad L(\Omega) = 1\,, \quad (A_n)_{n\in\mathbb{N}} \downarrow A \Longrightarrow \lim_{n\to\infty} L(A_n) = L(A) \tag{9}$$

and

$$(A_n)_{n\in\mathbb{N}} \uparrow A,\; A_n \text{ open}, n \in \mathbb{N},\; \Longrightarrow \lim_{n\to\infty} L(A_n) = L(A)\,. \tag{10}$$

Totally-monotone capacities were independently introduced to statistics by Strassen [34] and by Dempster (e.g. [16]), and form the basis of the Dempster-Shafer theory [33, 41]. Two-monotone capacities provide a superstructure upon the neighbourhood models commonly used in robust Neyman-Pearson testing [4, 28, 29].

Under quite mild regularity conditions, each two- or totally-monotone capacity $L(\cdot)$ leads to F-probability, and therefore to two- or totally-monotone C-probability, respectively.

Lemma 2 (After [29, Lemma 2.5])

Let $(\Omega, \mathcal{A})$ be a measurable space, where Ω is complete, separable and metrizable, and $\mathcal{A}$ is the corresponding Borel σ-field. If $L(\cdot)$ is a two- *or a* totally-monotone capacity, *and $U(\cdot) := 1 - L(\cdot^C)$ is the corresponding conjugated set-function, then $P(\cdot) := [L(\cdot); U(\cdot)]$ is C-probability, or totally monotone C-probability, respectively.*

In the next section $A_{(n)}$-based interval probability is derived, and it will be shown that this is totally-monotone C-probability, and therefore F-probability. Furthermore, the $A_{(n)}$-based $L(\cdot)$ is not a Choquet capacity, a fact that provides new theoretical insight into the relation between totally-monotone Choquet capacities and totally-monotone C-probabilities.

3 Predictive interval probability based on $A_{(n)}$

We start this section introducing predictive interval probabilities based on the assumption $A_{(n)}$, and present some basic properties. In subsection 3.1 we prove that these interval probabilities are F-probability. In subsection 3.2 we focus on the attractive property of local additivity which is important for conditioning as will be discussed in section 4. Finally, in subsection 3.3 we relate our inference to inner and outer measure and give a more detailed characterization of the structure arising.

Let $\mathcal{B}$ be the Borel σ-field over $I\!R$. For any element $B \in \mathcal{B}$, set-functions L and U for the event $X_{n+1} \in B$, based on data $x_1, \ldots, x_n$, which partition $I\!R$ into intervals $I_1, \ldots, I_{n+1}$, and the assumption $A_{(n)}$, are

$$L(X_{n+1} \in B) = \frac{1}{n+1}|\{j : I_j \subset B\}| \tag{11}$$

$$U(X_{n+1} \in B) = \frac{1}{n+1}|\{j : I_j \cap B \neq \emptyset\}|. \tag{12}$$

Throughout the paper, we leave the conditioning on data $x_1, \ldots, x_n$ out of the notation. $L(\cdot)$ and $U(\cdot)$ can be understood as bounds for the probability for the event $X_{n+1} \in B$, consistent with the probabilities as assessed by $A_{(n)}$. The lower bound $L(\cdot)$ is achieved by taking only probability mass into account that is necessarily within B, which is only the case for the probability mass $\frac{1}{n+1}$ per interval I_j if this interval is completely contained within B. The upper bound $U(\cdot)$ is achieved by taking all the probability mass into account that could possibly be within B, which is the case for the probability mass $\frac{1}{n+1}$ per interval I_j if the intersection of I_j and B is non-empty. Remark that, in this reasoning, we do allow positive probability masses in points. The way of deriving $L(\cdot)$ is similar in idea to the construction of belief functions. Notice, however, that $L(\cdot)$ cannot be a belief function, because the latter is defined for finite spaces only.

Next, we discuss some basic properties of $L(\cdot)$ and $U(\cdot)$. For all $B \in \mathcal{B}$, we have

$$\begin{aligned} L(X_{n+1} \in B) &= \frac{1}{n+1}|\{j : I_j \subset B\}| = \frac{1}{n+1}|\{j : I_j \cap B^c = \emptyset\}| \\ &= \frac{1}{n+1}\left[n+1-|\{j : I_j \cap B^c \neq \emptyset\}|\right] = 1 - U(X_{n+1} \in B). \end{aligned}$$

This means that $L(\cdot)$ and $U(\cdot)$ as defined in (11) and (12) satisfy (4). This conjugacy justifies that we mainly concentrate on $L(\cdot)$. Next we prove that total monotonicity holds for $L(\cdot)$. Let $B_i \in \mathcal{B}$, $i = 1, \ldots, q$, with finite integer $q \geq 2$ and let $S \subset \{1, \ldots, q\}$.

Lemma 3 (For total monotonicity)

$$(a) \quad \{j : I_j \subset \textstyle\bigcup_{i=1}^q B_i\} \supset \textstyle\bigcup_{i=1}^q \{j : I_j \subset B_i\}$$

$$(b) \quad \textstyle\bigcap_{i \in S}\{j : I_j \subset B_i\} = \{j : I_j \subset \textstyle\bigcap_{i \in S} B_i\}.$$

Proof

$$
\begin{aligned}
(a)\quad j_0 \in \bigcup_{i=1}^{q}\{j\,:\,I_j \subset B_i\} \quad &\Rightarrow \quad \exists_{i_o\in\{1,\ldots,q\}} : I_{j_0} \subset B_{i_0} \\
&\Rightarrow \quad I_{j_0} \subset \bigcup_{i=1}^{q} B_i \;\Rightarrow\; j_0 \in \{j\,:\,I_j \subset \bigcup_{i=1}^{q} B_i\} \\
(b)\quad j_0 \in \bigcap_{i\in S}\{j\,:\,I_j \subset B_i\} \quad &\Leftrightarrow \quad \forall_{i\in S} : I_{j_0} \subset B_i \\
&\Leftrightarrow \quad I_{j_0} \subset \textstyle\bigcap_{i\in S} B_i \;\Leftrightarrow\; j_0 \in \{j\,:\,I_j \subset \bigcap_{i\in S} B_i\}.
\end{aligned}
$$

It is easily seen that $\{j\,:\,I_j \subset \bigcup_{i=1}^{q} B_i\} \not\subset \bigcup_{i=1}^{q}\{j\,:\,I_j \subset B_i\}$. For example, observations $x_1 = 1$ and $x_2 = 2$ give $I_1 = (-\infty, 1)$, $I_2 = (1, 2)$ and $I_3 = (2, \infty)$. If we take $B_1 = [0.5, 1.5]$ and $B_2 = [1.5, 2.5]$, then $2 \in \{j\,:\,I_j \subset \bigcup_{i=1}^{2} B_i\}$, whereas $\bigcup_{i=1}^{2}\{j\,:\,I_j \subset B_i\} = \emptyset$.

Theorem 1 (Total monotonicity)
Set-function $L(\cdot)$, based on data $x_1, \ldots, x_n$ and the assumption $A_{(n)}$, as defined in (11), is totally monotone.

Proof
We prove that $L(\cdot)$ satisfies (8), using Lemma 3 and a standard inclusion-exclusion result from the theory of sets.

$$
\begin{aligned}
L(\bigcup_{i=1}^{q}\{X_{n+1} \in B_i\}) &= L(\{X_{n+1} \in \bigcup_{i=1}^{q} B_i\}) \;=\; \frac{1}{n+1}|\{j\,:\,I_j \subset \bigcup_{i=1}^{q} B_i\}| \\
&\geq \frac{1}{n+1}|\bigcup_{i=1}^{q}\{j\,:\,I_j \subset B_i\}| \\
&= \frac{1}{n+1}\sum_{\emptyset\neq S\subset\{1,\ldots,q\}} (-1)^{|S|+1}|\bigcap_{i\in S}\{j\,:\,I_j \subset B_i\}| \\
&= \frac{1}{n+1}\sum_{\emptyset\neq S\subset\{1,\ldots,q\}} (-1)^{|S|+1}|\{j\,:\,I_j \subset \bigcap_{i\in S} B_i\}| \\
&= \sum_{\emptyset\neq S\subset\{1,\ldots,q\}} (-1)^{|S|+1} L(\{X_{n+1} \in \bigcap_{i\in S} B_i\}) \\
&= \sum_{\emptyset\neq S\subset\{1,\ldots,q\}} (-1)^{|S|+1} L(\bigcap_{i\in S}\{X_{n+1} \in B_i\}).
\end{aligned}
$$

Total-monotonicity leads to (so-to-say static) coherence in Walley's [36] sense. (For a proof use the fact that total-monotonicity implies two-monotonicity and apply [35, Corollary 6.3].) According to Walley's generalized betting interpretation for interval probability

[36, Chapter 2], this means that if we are acting according to our interval probabilities, nobody can place a Dutch book against us at any fixed moment in time.

3.1 $A_{(n)}$-based interval probability is F-probability

While Walley's concept of coherence is strictly connected with finite additivity of the probabilities considered [36, Chapter 3.3], we now turn to a stronger property of internal consistency, which is based on countable additivity of classical probability. We will embed our $A_{(n)}$-based interval probability into Weichselberger's system of interval probability by showing that it is F-probability. A natural way to start would be to check the conditions (9) and (10) and then to apply Lemma 2. But a simple counter example shows that this way of proceeding is not applicable. The set-function $L(\cdot)$ derived from (11) does not satisfy (10).

Example 1 (Counter example continuity)
Consider the set-function $L(\cdot)$ for the random quantity X_3, based on two observations, $x_1 = \frac{1}{2}$ and $x_2 = 1$, and the assumption $A_{(2)}$. Define the following sequence of open[1] Borel sets: $A_i = (0, 1 - \frac{1}{i})$, with $A = \bigcup_{i=1}^{\infty} A_i = (0,1)$. It is clear that $L(X_3 \in A_i) = 0$ for every $i \geq 1$, so

$$\lim_{s\to\infty} L(X_3 \in \bigcup_{i=1}^{s} A_i) \;=\; \lim_{s\to\infty} L(X_3 \in A_s) \;=\; 0,$$

but $L(X_3 \in A) = \frac{1}{3}$. Hence, this set-function $L(\cdot)$ does not satisfy (10).

Note further that in this counter example the limit A is different from $\Omega = I\!R$, so $A^C \neq \emptyset$. This makes impossible the use of Buja's ([8]) elegant method of applying the Kuratowski isomorphism theorem to establish (10) in a non-standard topology, which has been quite successful for proving (10) for the usual neighbourhood models arising in robust statistics.

Nevertheless our interval probabilities are F-probabilities. This can be deduced from the following theorem, which also establishes a weakened form of continuity and prepares the ground for a simplification of the practical calculation of conditional probabilities.

Theorem 2
Let $A_1 \subset A_2 \subset \ldots \subset A_i \subset \ldots$ be an infinite sequence $\mathcal{S}$ of Borel sets ($A_i \in \mathcal{B}$ for all $i \geq 1$)

[1] By embedding the situation into the product-space $(I\!R^{n+1}, \mathcal{B}^{n+1})$ endowed with the standard topology and taking X_{n+1} as the $(n+1)st$-projection, the set $\{X_{n+1} \in A\}$ is open iff A is open.

with $A := \bigcup_{i=1}^{\infty} A_i$ *satisfying*

Assumption 1: For each j with $I_j \subset A$ there is an i_j such that $I_j \subset A_{i_j}$.

Then there exists classical probability $p_{\mathcal{S}}(.)$ *with*

$$
\begin{aligned}
p_{\mathcal{S}}(X_{n+1} \in B) &\geq L(X_{n+1} \in B)\,, \quad \forall B \in \mathcal{B}\,, && (13)\\
p_{\mathcal{S}}(X_{n+1} \in A_i) &= L(X_{n+1} \in A_i)\,, \quad \forall A_i \in \mathcal{S}\,, && (14)\\
p_{\mathcal{S}}(X_{n+1} \in A) &= L(X_{n+1} \in A)\,. && (15)
\end{aligned}
$$

Proof

First we define a set $\{a_1, \ldots, a_j \ldots, a_{n+1}\}$ of real values with $a_j \in I_j$, $j = 1, \ldots, n+1$. For every j the following method is used to define a corresponding a_j.

- If $I_j \cap A^C \neq \emptyset$, then take as a_j any element of $I_j \cap A^C$.
- In the alternative case, $I_j \subset A$, let a_j be any element of $I_j \cap \left(A_{n_j} \setminus A_{n_j - 1}\right)$, where $n_j := \min\{n \in I\!N : I_j \subset A_n\}$. (Because of Assumption 1 the number n_j is well defined.)

Now we put probability mass $\frac{1}{n+1}$ into every point a_j and consider the set-function $p_{\mathcal{S}}(\cdot)$ defined by

$$p_{\mathcal{S}}(X_{n+1} \in B) := \frac{1}{n+1} \sum_{j=1}^{n+1} 1_B(a_j), \quad \forall B \in \mathcal{B}\,,$$

where $1_B(\cdot)$ denotes the indicator function. Since $a_j \in I_j$ for all j, we have for every $B \in \mathcal{B}$ the relation $I_j \subset B \Rightarrow a_j \in B$, and therefore

$$
\begin{aligned}
L(X_{n+1} \in B) = \frac{1}{n+1} \left|\{j : I_j \subset B\}\right| &\leq \frac{1}{n+1} \left|\{j : a_j \in B\}\right| \\
&= \frac{1}{n+1} \sum_{j=1}^{n+1} 1_B(a_j) = p_{\mathcal{S}}(X_{n+1} \in B)\,,
\end{aligned}
$$

as stated in (13).

This construction ensures that for every j and every member A_i of the sequence $\mathcal{S}$ the relation $a_j \in A_i$ is only possible if $i \geq n_j$, which implies $I_j \subset A_i$. Therefore

$$
\begin{aligned}
p_{\mathcal{S}}(X_{n+1} \in A_i) &= \frac{1}{n+1} \sum_{j=1}^{n+1} 1_{A_i}(a_j) = \frac{1}{n+1} \left|\{j : a_j \in A_i\}\right| \\
&\leq \frac{1}{n+1} \left|\{j : I_j \subset A_i\}\right| = L(X_{n+1} \in A_i)\,,
\end{aligned}
$$

leading together with (13) to (14). And, analogously, the fact that $a_j \in A$ can only be true if $I_j \subset A$ holds, gives us (15).

To complete the proof we have to check that $p_{\mathcal{S}}(\cdot)$ satisfies Kolmogorov's axioms. Obviously $p_{\mathcal{S}}(\cdot)$ is nonnegative with $p_{\mathcal{S}}(X_{n+1} \in \mathbb{R}) = 1$. For countable additivity of $p_{\mathcal{S}}(\cdot)$, note its finite additivity as a first step. For two disjoint Borel sets B_1 and B_2 we have

$$\begin{aligned} p_{\mathcal{S}}(X_{n+1} \in B_1 \cup B_2) &= \frac{1}{n+1} \sum_{j=1}^{n+1} 1_{B_1 \cup B_2}(a_j) \\ &= \frac{1}{n+1} \sum_{j=1}^{n+1} (1_{B_1}(a_j) + 1_{B_2}(a_j)) \\ &= p_{\mathcal{S}}(X_{n+1} \in B_1) + p_{\mathcal{S}}(X_{n+1} \in B_2) \,. \end{aligned}$$

For an arbitrary sequence of pairwise disjoint sets $B_i \in \mathcal{B}$, $i \in \mathbb{N}$, we note that there can only be a *finite* collection of these sets which have non-empty intersection with the set $\{a_j,\ j = 1, \ldots, n+1\}$. This means that there exists a finite set $J \subset \mathbb{N}$ such that

$$\frac{1}{n+1} \sum_{j=1}^{n+1} 1_{\bigcup\limits_{i=1}^{\infty} B_i}(a_j) = \frac{1}{n+1} \sum_{j=1}^{n+1} 1_{\bigcup\limits_{i \in J} B_i}(a_j)$$

and $\frac{1}{n+1} \sum_{j=1}^{n+1} 1_{B_i}(a_j) = 0$ for all $i \in \mathbb{N} \setminus J$. Then

$$\begin{aligned} p_{\mathcal{S}}\left(X_{n+1} \in \bigcup_{i=1}^{\infty} B_i\right) &= \frac{1}{n+1} \sum_{j=1}^{n+1} 1_{\bigcup\limits_{i=1}^{\infty} B_i}(a_j) &&= \frac{1}{n+1} \sum_{j=1}^{n+1} 1_{\bigcup\limits_{i \in J} B_i}(a_j) \\ &= \frac{1}{n+1} \sum_{j=1}^{n+1} \left(\sum_{i \in J} 1_{B_i}(a_j) \right) &&= \frac{1}{n+1} \sum_{i \in J} \left(\sum_{j=1}^{n+1} 1_{B_i}(a_j) \right) \\ &= \frac{1}{n+1} \sum_{i \in J} p_{\mathcal{S}}(X_{n+1} \in B_i) + 0 &&= \frac{1}{n+1} \sum_{i=1}^{\infty} p_{\mathcal{S}}(X_{n+1} \in B_i) \,. \end{aligned}$$

This completes the proof of Theorem 2, from which we will deduce several corollaries. One of these is the main result of this section, namely that our method leads to F-probability.

Corollary 1 (F-probability)
$(\mathbb{R}, \mathcal{B}, L(.))$, with set-function $L(\cdot)$ based on data $x_1, \ldots, x_n$ and the assumption $A_{(n)}$, is a totally-monotone C-probability-field, and therefore an F-probability field.

Proof
The total-monotonicity is already stated in Theorem 1. To prove the F-property, apply

for each $B \in \mathcal{B}$ Theorem 2 to the degenerate sequence $A_i = B, i \geq 1$ and $A = B$. Together with the conjugacy property of $L(\cdot)$ the relation (13) shows that there is a non-empty structure; the F-property (3) is given by (14), again with the help of conjugacy.

A weakened version of (10) can also be achieved. Assumption 1 prevents situations as in the counter example above.

Corollary 2 (Pseudo-continuity)
For all sequences $\mathcal{S}$, as introduced in Theorem 2, and for which Assumption 1 holds, we have

$$\lim_{s\to\infty} L(X_{n+1} \in \bigcup_{i=1}^{s} A_i) \;=\; L(X_{n+1} \in \lim_{s\to\infty} \bigcup_{i=1}^{s} A_i).$$

Proof
This follows from (14) and (15) in Theorem 2 and continuity of classical probabilities.

3.2 Local additivity

Before considering at conditioning and updating in the next sections, we briefly discuss some further properties of the unconditional interval probabilities connected with the special nature of $A_{(n)}$. In particular we use the fact that probability statements on intervals I_j and unions of them are precise and that these sets provide a natural partition of the sample space.

Definition 4 (Natural partition)
A finite collection $\mathcal{C} = \{C_1, \ldots, C_q, \ldots C_l\}$ of events is called a natural partition, *if there are non-empty pairwise disjoint index sets $\mathcal{J}_q \subset \{1, \ldots, n+1\}$, $q = 1, \ldots, l$, with $\cup_{q=1}^{l} \mathcal{J}_q = \{1, \ldots, n+1\}$, such that $C_q = \cup_{j\in\mathcal{J}_q} I_j$, $q = 1, \ldots, l$.*

Interval probabilities are typically non-additive. Our set-functions, however, are 'locally additive' with respect to natural partitions. This property is important for several results stated below and for practical applications.

Theorem 3 (Local additivity)
Let $\mathcal{C} = \{C_1, \ldots, C_q, \ldots C_l\}$ be a natural partition. Then for every event $A \in \mathcal{B}$

$$L(X_{n+1} \in A) = \sum_{q=1}^{l} L(X_{n+1} \in A \cap C_q) \quad \textit{and} \quad U(X_{n+1} \in A) = \sum_{q=1}^{l} U(X_{n+1} \in A \cap C_q)\,. \quad (16)$$

Proof:

$$\begin{aligned}\sum_{q=1}^{l} L(X_{n+1} \in A \cap C_q) &= \frac{1}{n+1}\sum_{q=1}^{l} |j : I_j \subset A \cap C_q| = \frac{1}{n+1}\sum_{q=1}^{l} |j \in \mathcal{J}_q : I_j \subset A| \\ &= \frac{1}{n+1}|j : I_j \subset A| = L(X_{n+1} \in A)\,,\end{aligned}$$

and analogously for $U(\cdot)$.

3.3 Further characterizations

We conclude this section by looking at the relation of these inferences to inner and outer measure and by giving a characterization of the structure of the F-probability field arising.

Consider the classical probability field $(I\!R, \mathcal{D}, p_{\mathcal{D}}(\cdot))$, where the σ-field $\mathcal{D}$ is the power set of $\{I_1, \ldots, I_{n+1}\}$, and

$$p_{\mathcal{D}}(D) := \frac{|\{j : I_j \subset D|\}}{n+1}, \quad D \in \mathcal{D}\,.$$

Then the set-functions $L(\cdot)$ and $U(\cdot)$ from (11) and (12) can be reformulated as

$$L(X_{n+1} \in A) = \sup_{D \in \mathcal{D},\, D \subset A} p_{\mathcal{D}}(D) \;\text{ and }\; U(X_{n+1} \in A) = \inf_{D \in \mathcal{D},\, D \supset A} p_{\mathcal{D}}(D), \quad A \in \mathcal{B}\,.$$

This means that $L(\cdot)$ and $U(\cdot)$ can equivalently be interpreted as the extension of the classical probability induced by the assumption $A_{(n)}$ to the whole Borel σ-field via inner and outer measure.

Walley [36, p.125f.] relates inner and outer measure to coherent interval probability and a set of finitely additive classical probabilities. Using σ-additive probability as a primitive we arrive at results which are similar in the end. The structure of the F-probability field resulting from our inferences can be characterized in the following way.

Theorem 4
The structure $\mathcal{M}$ of the F-probability field $(\Omega; \mathcal{A}; L(\cdot))$, with $L(\cdot)$ and $U(\cdot)$ arising from $A_{(n)}$-based inference with data $x_1, \ldots, x_n$, as defined in (11) and (12), has the form

$$\mathcal{M} = \left\{ p(\cdot) \in \mathcal{K}\,(I\!R, \mathcal{B}) \;\middle|\; p(I_j) = \frac{1}{n+1}, \quad \forall j = 1, \ldots, n+1 \right\}. \tag{17}$$

Proof:
Let $\mathcal{V}_0$ be the right hand side of (17). If we compare this with the structure

$$\mathcal{M} = \{p(\cdot) \in \mathcal{K}\,(I\!R, \mathcal{B}) \mid L(X_{n+1} \in A) \leq p(X_{n+1} \in A) \leq U(X_{n+1} \in A), \quad \forall A \in \mathcal{B}\}$$

of the F-probability field $(I\!R; \mathcal{B}; L(\cdot))$, then we immediately arrive through $L(X_{n+1} \in I_j) = U(X_{n+1} \in I_j) = \frac{1}{n+1}$, $j = 1, \ldots, n+1$, at the fact that $\mathcal{M}$ contains all the restrictions in $\mathcal{V}_0$ and some more. Therefore $\mathcal{M} \subset \mathcal{V}_0$.

To prove that $\mathcal{M} \supset \mathcal{V}_0$, consider the F-probability field $(I\!R; \mathcal{B}; L_0(\cdot))$ deduced from $\mathcal{V}_0$ via (6) and prove the equivalent relation $L(\cdot) \leq L_0(\cdot)$. This is clear for every $B \subset I_{j_0}$, for some $j_0 \in \{1, \ldots, n+1\}$, because $L(X_{n+1} \in B) = 0 \leq L_0(X_{n+1} \in B)$ if $B \neq I_{j_0}$, and $L(X_{n+1} \in B) = \frac{1}{n+1} = L_0(X_{n+1} \in B)$ if $B = I_{j_0}$. Arbitrary $B \in \mathcal{B}$ can be partitioned into $B = \cup_{j=1}^{n+1} B \cap I_j$. From above and from the local additivity in Theorem 3 we obtain, using superadditivity (cf. (5)) of $L_0(\cdot)$ in the last inequality,

$$L(X_{n+1} \in B) = \sum_{j=1}^{n+1} L(X_{n+1} \in B \cap I_j) \leq \sum_{j=1}^{n+1} L_0(X_{n+1} \in B \cap I_j) \leq L_0(X_{n+1} \in B)\,.$$

Denoting the structure of $(I\!R; \mathcal{B}; L_0(\cdot))$ by $\mathcal{M}_0$, this leads to $\mathcal{V}_0 \subset \mathcal{M}_0 \subset \mathcal{M}$, which completes the proof.

We have shown that the structure of the F-probability field obtained by $A_{(n)}$-based inference consists of all classical probabilities being in accordance with $A_{(n)}$ on the intervals I_j. This emphasizes that the inferences described here are only exploiting the assumption $A_{(n)}$ in a perfect manner without adding any further assumptions.

4 Conditioning

In this section we consider conditioning on an event $\{X_{n+1} \in C\}$, $C \in \mathcal{B}$, within our framework of predictive inference for X_{n+1}, based on observations $x_1, \ldots, x_n$ and the assumption $A_{(n)}$. After some general remarks on different ways of conditioning with interval probability, we derive a theorem of total probability for natural partitions in subsection 4.1. In this subsection we also show that, for conditioning on $X_{n+1} \in C$, where C is an element of a natural partition, two ways of using this information coincide. One can either condition on this event, or consider a conditional form of $A_{(n)}$. Conditioning on arbitrary events is presented in subsection 4.2, where we derive convenient expressions from a computational point of view.

For a long time there has been confusion between mainly two competing approaches to conditional interval probability. The first one is called the *intuitive concept* in this paper, following Weichselberger [38, 40]. It concentrates on the structure and defines conditional interval probability via the infimum and supremum over all classical conditional probabilities formed by elements of the structure. Given an F-probability field $\mathcal{F} = (\Omega; \mathcal{A}; L(\cdot))$ with structure $\mathcal{M}$,

$$L^i(A|C) := \inf_{p \in \mathcal{M}} p(A|C) \;\text{ and }\; U^i(A|C) := \sup_{p \in \mathcal{M}} p(A|C), \quad \forall A, C \in \mathcal{A},\ L(C) > 0. \tag{18}$$

By construction, $L^i(\cdot|C)$ and $U^i(\cdot|C)$ are conjugated (relation (4)). Furthermore, for every suitable C the triple $(C; \mathcal{A}_C; L^i(\cdot|C))$ with $\mathcal{A}_C := \{A \cap C | A \in \mathcal{A}\}$ is an F-probability field.

This can be seen by applying (6) in Lemma 1 to the set $\{p(\cdot \cap C)/p(C) \mid p(\cdot) \in \mathcal{M}\}$. For example, this concept has been propagated by Walley [36, chapter 6], who derives it from coherence considerations between unconditional and conditional (contingent) gambles.

Another definition of conditional interval probability dates back to Dempster [16] and his proposed method of statistical inference. Nowadays this is often used independently of its original motivation as *Dempster's rule of conditioning*, mainly in the area of artificial intelligence. The upper interval limit for $A|C$ is defined as the ratio of the upper interval limits for $A \cap C$ and C, the lower one by ensuring conjugacy (4),

$$U^D(A \mid C) := \frac{U(A \cap C)}{U(C)} \quad \text{and} \quad L^D(A \mid C) := 1 - \frac{U(A^C \cap C)}{U(C)}, \qquad \forall A, C \in \mathcal{A},\ U(C) > 0. \tag{19}$$

In the case of totally-monotone C-probability, $[L^D(A \mid C); U^D(A \mid C)]$ is also totally-monotone C-probability, but the operation is not closed under F-probability. This concept of conditional interval probability has experienced many modifications, see [42] for a comparison of different proposals.

Recently, the idea has gained more and more acceptance that there cannot be one single concept of conditional interval probability which is appropriate for all purposes. It is becoming clear that there should be a coexistence of several different concepts, based on different tasks of conditional probability which coincide in the case of classical probability. Dubois and Prade [22] use the intuitive concept for what they call 'focusing', and Dempster's rule for 'conditioning'. In Weichselberger's theory [38, 40] the intuitive concept is supplemented by the 'canonical concept'[2], based on a canon of desirable properties.

As already mentioned, all definitions are generalizations of classical conditional probability. In the special case of interval probability where $L(C) = U(C)$ for the event C on which is conditioned, all these concepts coincide. We then write $L(\cdot \mid C)$ and $U(\cdot \mid C)$ for the conditional probability.

4.1 Conditioning on an element of a natural partition

We now focus on conditioning on $X_{n+1} \in C$, where C is an element of a natural partition, in the sense of Definition 4, so $C = \cup_{j \in \mathcal{J}_C} I_j$ for an appropriate set $\mathcal{J}_C \subset \{1, \ldots, n+1\}$. Then $L(X_{n+1} \in C) = U(X_{n+1} \in C)$, and we have the special case mentioned above, arriving at a type of conditional interval probability which is a synthesis of the different concepts, bringing together all their nice properties. For every $A \in \mathcal{B}$,

$$L(X_{n+1} \in A | X_{n+1} \in C) \quad = \quad \inf_{p(\cdot) \in \mathcal{M}} \frac{p(X_{n+1} \in A \cap C)}{p(X_{n+1} \in C)}$$

[2] We do not define the canonical concept here because we will only use it implicitly in a special case.

$$= \frac{L(X_{n+1} \in A \cap C)}{L(X_{n+1} \in C)} = \frac{L(X_{n+1} \in A \cap C)}{U(X_{n+1} \in C)} \tag{20}$$

$$= \frac{\frac{1}{n+1}\,|\{j|I_j \subset A \cap C\}|}{\frac{1}{n+1}\,|\{j|I_j \subset C\}|} = \frac{|\{j \in \mathcal{J}_\mathcal{C}|I_j \subset A\}|}{|\mathcal{J}_\mathcal{C}|},$$

which agrees with (19) because

$$\frac{|\{j \in \mathcal{J}_\mathcal{C}|I_j \subset A\}|}{|\mathcal{J}_\mathcal{C}|} = 1 - \frac{U(X_{n+1} \in A^c \cap C)}{U(X_{n+1} \in C)}.$$

And analogously

$$\begin{aligned} U(X_{n+1} \in A|X_{n+1} \in C) &= \sup_{p(\cdot)\in\mathcal{M}} \frac{p(X_{n+1} \in A \cap C)}{p(X_{n+1} \in C)} \\ &= \frac{U(X_{n+1} \in A \cap C)}{U(X_{n+1} \in C)} = \frac{U(X_{n+1} \in A \cap C)}{L(X_{n+1} \in C)} \qquad (21) \\ &= \frac{\frac{1}{n+1}\,|\{j|I_j \cap A \cap C \neq \emptyset\}|}{\frac{1}{n+1}\,|\{j|I_j \cap C \neq \emptyset\}|} = \frac{|\{j \in \mathcal{J}_\mathcal{C}|I_j \cap A \neq \emptyset\}|}{|\mathcal{J}_\mathcal{C}|}. \end{aligned}$$

It should be noted that this conditional interval probability is perfectly consistent with another way of exploiting the information $X_{n+1} \in C$, namely a so-to-say *conditional* $A_{(n)}$ inference. Knowing $X_{n+1} \in C = \cup_{j\in\mathcal{J}_\mathcal{C}} I_j$, and assuming nothing else, we would divide the probability mass 1 among the $|\mathcal{J}_\mathcal{C}|$ intervals I_j, $j \in \mathcal{J}_\mathcal{C}$. Reasoning in accordance to (11) and (12) would lead to

$$L_C(X_{n+1} \in A) = \frac{|\{j \in \mathcal{J}_\mathcal{C}|I_j \subset A\}|}{|\mathcal{J}_\mathcal{C}|}$$

and

$$U_C(X_{n+1} \in A) = \frac{|\{j \in \mathcal{J}_\mathcal{C}|I_j \cap A \neq \emptyset\}|}{|\mathcal{J}_\mathcal{C}|},$$

which are equal to $L(X_{n+1} \in A|X_{n+1} \in C)$ and $U(X_{n+1} \in A|X_{n+1} \in C)$ from above. Therefore, starting with $A_{(n)}$, deriving interval probability from it and then conditioning on the information $X_{n+1} \in C = \cup_{j\in\mathcal{J}_\mathcal{C}} I_j$ gives the same result as doing the conditioning process first and then calculating interval probability from that conditional $A_{(n)}$.

For natural partitions it can easily be shown that the theorem of total probability is valid for our nonparametric predictive inference.

Theorem 5 (Theorem of total probability for natural partitions)
Let $\mathcal{C} = \{C_1, \ldots, C_q, \ldots C_l\}$ *be a* natural partition. *Then the set-functions* $L(\cdot)$ *and* $U(\cdot)$*,*

based on the assumption $A_{(n)}$ and data $x_1, \ldots, x_n$, and the corresponding conditional set-functions, satisfy

$$\begin{aligned} L(X_{n+1} \in A) &= \sum_{q=1}^{l} L(X_{n+1} \in A | X_{n+1} \in C_q) \cdot L(X_{n+1} \in C_q)\,, \quad \forall A \in \mathcal{B}, \\ U(X_{n+1} \in A) &= \sum_{q=1}^{l} U(X_{n+1} \in A | X_{n+1} \in C_q) \cdot U(X_{n+1} \in C_q)\,, \quad \forall A \in \mathcal{B}. \end{aligned}$$

Proof:

This is an immediate consequence of Theorem 3. For example,

$$\begin{aligned} L(X_{n+1} \in A) &= \sum_{q=1}^{l} L(X_{n+1} \in A \cap C_q) \\ &= \sum_{q=1}^{l} L(X_{n+1} \in A \cap C_q) \cdot \left(\frac{|\mathcal{J}_q|}{n+1}\right)^{-1} \cdot \frac{|\mathcal{J}_q|}{n+1} \\ &= \sum_{q=1}^{l} L(X_{n+1} \in A | X_{n+1} \in C_q) \cdot L(X_{n+1} \in C_q)\,. \end{aligned}$$

Conditioning on elements of the partition consisting of only a single interval I_j, leads to

$$L(X_{n+1} \in A | X_{n+1} \in I_j) = 0 \text{ and } U(X_{n+1} \in A | X_{n+1} \in I_j) = 1\,, \quad \forall A \in \mathcal{B}\,,$$

which is another paraphrase of the fact that we do not add further assumptions to $A_{(n)}$. We can only be precise concerning the whole intervals I_j, there is uncertainty in the strict sense about what is happening inside these I_j.

4.2 Conditioning on arbitrary events

When considering conditional interval probability for events of the form $X_{n+1} \in A \mid X_{n+1} \in C$, where C does not consist of (a union of) intervals I_j, the intuitive concept of conditional interval probability according to (18) seems to be the appropriate choice, since we are interested in the consequences which the structure implies for conditional events. For practical application Corollary 4 is of importance, because it allows us to circumvent the complex optimisation underlying the original definition in (18). The proof requires one more result which follows easily from Theorem 2, by defining $A_1 = B_1$ and $A_i = B_2^c$, for $i \geq 2$.

Corollary 3

For each $B_1, B_2 \in \mathcal{B}$ with $B_2 \subset B_1^c$ there is a $p_{B_1,B_2}(\cdot) \in \mathcal{M}$ with $p_{B_1,B_2}(X_{n+1} \in B_1) = L(X_{n+1} \in B_1)$ and $p_{B_1,B_2}(X_{n+1} \in B_2) = U(X_{n+1} \in B_2)$.

Corollary 4 (Conditional interval probability)
For every $A, C \in \mathcal{B}$ with $L(X_{n+1} \in C) > 0$ we have

$$L^i(X_{n+1} \in A | X_{n+1} \in C) = \frac{L(X_{n+1} \in A \cap C)}{L(X_{n+1} \in A \cap C) + U(X_{n+1} \in A^c \cap C)}$$

and

$$U^i(X_{n+1} \in A | X_{n+1} \in C) = \frac{U(X_{n+1} \in A \cap C)}{U(X_{n+1} \in A \cap C) + L(X_{n+1} \in A^c \cap C)}.$$

Proof
This follows from Corollary 3 and the fact that $(A^c \cap C) \subset (A \cap C)^c$, using the relation

$$\frac{p(X_{n+1} \in A \cap C)}{p(X_{n+1} \in C)} = \frac{p(X_{n+1} \in A \cap C)}{p(X_{n+1} \in A \cap C) + p(X_{n+1} \in A^C \cap C)} = \frac{1}{1 + \frac{p(X_{n+1} \in A^C \cap C)}{p(X_{n+1} \in A \cap C)}}.$$

A slight generalization of Corollary 4 can be achieved by relaxing in (18) the condition on C to $U(X_{n+1} \in C) > 0$ and by replacing $p \in \mathcal{M}$ by $p \in \mathcal{M}$, $p(X_{n+1} \in C) > 0$.

When appropriate, also Dempster's rule of conditioning (cf. (19)) can be powerfully applied to our nonparametric predictive inferences. The resulting conditional interval probability $[L^D(\cdot \mid C), U^D(\cdot \mid C)]$ is again totally-monotone C-probability, and therefore F-probability. To see the total monotonicity, use Theorem 1, formulate relation (8) in terms of $U(\cdot)$, and note that the inequality is not changed by dividing through by $U(X_{n+1} \in C) > 0$. The property of being F-probability follows from Theorem 2 by setting, for every $A \in \mathcal{B}$, $A_1 = C^C$ and $A_i = (A \cap C)^C$ for all $i \geq 2$.

5 Updating

In this section we discuss how extra observations affect our nonparametric predictive inference. We analyse inference on X_{n+2}, both before and after the value of X_{n+1} becomes available. In subsection 5.1 we present the rather straightforward mechanism of updating on the basis of a new observation. Internal consistency of such updating is shown in subsection 5.2.

The assumption $A_{(n)}$ gives a predictive probability for a future observation on the basis of data $x_1, \ldots, x_n$. Although our focus has so far been on X_{n+1}, our results hold for any single future observation X_{n+l}, $l \geq 1$. However, care should be taken that such future observations are not independent (conditionally on the n observations [11, 15, 25]).

On the basis of the first n observations, assuming $A_{(n)}$, the predictive inferences for X_{n+2} are again derived from the assessments

$$p(X_{n+2} \in I_j) = \frac{1}{n+1}, \text{ for } j = 1, \ldots, n+1.$$

Thus, interval probabilities for $X_{n+2} \in B$, $B \in \mathcal{B}$, are derived as presented before for X_{n+1}, and the same properties hold again.

5.1 Basic considerations of updating

When considering updating, we are interested in the effect of additional observations on our inferences, in contrast to conditioning where we exploit information on the same random quantity. Therefore, in section 4, where we were concerned with conditioning, we looked at interval probabilities for $X_{n+1} \in A \,|\, X_{n+1} \in C$, while now we study interval probabilities for X_{n+2} given the concrete value of X_{n+1}.

If the value of X_{n+1} becomes available, say $X_{n+1} = d$, then we simply obtain updated inferences for X_{n+2} by adding d to the observations $x_1, \ldots, x_n$. Now we assume $A_{(n+1)}$, for which we in effect maintain the same intervals I_j as before, with the exception that the interval in which d falls, say I_{j_d}, is divided into two intervals, (x_{j_d-1}, d) and (d, x_{j_d}), and the probability mass assigned to each interval is $\frac{1}{n+2}$. Interval probabilities for X_{n+2} to be in any $B \in \mathcal{B}$ are derived as before, and all the results remain valid, *mutatis mutandis*. Updating turns out to be a straightforward exercise in our approach, as it is always dealt with through replacing $A_{(n)}$ by $A_{(n+1)}$, or $A_{(n+l)}$ if $l \geq 1$ further observations have become available.

5.2 Internal consistency of updating

We now consider the relation between predictive inferences for X_{n+2} before and after observation of X_{n+1}, and we show a strong consistency of our procedure of updating.

If we are interested in predictive inferences for X_{n+2}, based on observations $x_1, \ldots, x_n$, and assuming $A_{(n)}$, then we can derive these in two ways. First, in a direct manner as described above, leading to

$$L(X_{n+2} \in B) = \frac{1}{n+1} |\{j : I_j \subset B\}|$$

and

$$U(X_{n+2} \in B) = \frac{1}{n+1} |\{j : I_j \cap B \neq \emptyset\}|,$$

for $B \in \mathcal{B}$.

Secondly, we can include in our reasoning the as yet unknown X_{n+1}, via the theorem of total probability. The only inference about X_{n+1} that we can make without additional assumptions is $p(X_{n+1} \in I_j) = \frac{1}{n+1}$, for all $j = 1, \ldots, n+1$. Bringing X_{n+1} into the argument should not change the result, and motivates consideration of set-function $\tilde{L}(\cdot)$ for X_{n+2}, based on $x_1, \ldots, x_n$, with $B \in \mathcal{B}$,

$$\tilde{L}(X_{n+2} \in B) := \sum_{j=1}^{n+1} L(X_{n+2} \in B \mid X_{n+1} \in I_j)\, p(X_{n+1} \in I_j). \tag{22}$$

Under a quite weak assumption on B, this is indeed equivalent to considering X_{n+2} directly. Of course, this additionally requires the natural assumption that one will be happy assuming $A_{(n+1)}$ as the basis for further inferences, after getting to know the value of X_{n+1}.

Theorem 6 (Internal consistency of updating)
Set-functions $L(\cdot)$ and $\tilde{L}(\cdot)$, describing inference based on data $x_1, \ldots, x_n$ and assumption $A_{(n)}$, as defined in (11) and (22), respectively, are equivalent in the following sense.

$$\tilde{L}(X_{n+2} \in B) = L(X_{n+2} \in B),$$

for all $B \in \mathcal{B}$ with $|\{I_j \cap B^c\}| \neq 1$ for all $j = 1, \ldots, n+1$,

Proof:
The main argument is as follows, the use of the assumption $|\{I_j \cap B^c\}| \neq 1$ for all $j = 1, \ldots, n+1$ for the second equality is explained below.

$$\begin{aligned}
\tilde{L}(X_{n+2} \in B) &= \sum_{r=1}^{n+1} L(X_{n+2} \in B \mid X_{n+1} \in I_r)\, p(X_{n+1} \in I_r) \\
&= \frac{1}{n+1} \sum_{r=1}^{n+1} \left(\frac{1}{n+2} \left[|\{j \,:\, I_j \subset B\}| + 1_{\{I_r \subset B\}} \right] \right) \\
&= \frac{1}{(n+1)(n+2)} \left((n+1)|\{j \,:\, I_j \subset B\}| + |\{j \,:\, I_j \subset B\}| \right) \\
&= \frac{1}{(n+1)} |\{j \,:\, I_j \subset B\}| \\
&= L(X_{n+2} \in B).
\end{aligned}$$

Detailed study of $L(X_{n+2} \in B \mid X_{n+1} \in I_{j_d})$ reveals the exact nature of the relation between the set-functions for X_{n+2} with and without conditioning on the as yet unknown value of X_{n+1}. It is easily derived that

$$L(X_{n+2} \in B \,|\, X_{n+1} \in I_{j_d}) \;=\; \begin{cases} \frac{1}{n+2}|\{j\,:\,I_j \subset B\}| & \text{if } I_{j_d} \cap B = \emptyset \\ \frac{1}{n+2}\left[|\{j\,:\,I_j \subset B\}| + 1\right] & \text{if } I_{j_d} \subset B \\ \frac{1}{n+2}\left[|\{j\,:\,I_j \subset B\}| + \underline{A}\right] & \text{if } I_{j_d} \cap B \neq \emptyset \\ & \quad \text{and } I_{j_d} \cap B^c \neq \emptyset \end{cases}$$

where the term $\underline{A}$ needs further explanation. Using the notation $I_{j_d} = (x_{j_d-1}, x_{j_d})$, as before, the term $\underline{A}$ is, without any further assumptions, the maximum achievable lower bound for

$$A \;=\; 1_{\{(x_{j_d-1},X_{n+1})\subset B\}} + 1_{\{(X_{n+1},x_{j_d})\subset B\}},$$

which is the relevant term given $X_{n+1} \in I_{j_d}$ and the conditions $I_{j_d} \cap B \neq \emptyset$ and $I_{j_d} \cap B^c \neq \emptyset$. It is easily seen that $\underline{A} = 0$ in all cases, except if $|\{I_{j_d} \cap B^c\}| = 1$, in which case $\underline{A} = 1$. This special case is the situation where only one singleton within I_{j_d} is not in B, and therefore splitting I_{j_d} gives necessarily at least one new interval that is completely contained within B.

Similar detailed considerations lead to analogous results for $U(\cdot)$, showing that not only the static part, based on conditioning, of our nonparametric predictive inference is consistent, but also the dynamic aspect of updating in the light of new observations is internally consistent.

6 Concluding Remarks

This paper presents properties of nonparametric predictive inference based on $A_{(n)}$, focussing on its position in the theory of interval probability. Strong internal consistency results have been derived. In particular, the property of F-probability has been proved for our interval probabilities, and for the corresponding conditional interval probabilities according to both the intuitive concept as well as Dempster's rule. Consistency in updating has also been shown.

These results support the theoretical position of our inferential method. Its practical value depends more strongly on possible applications. In the introduction, we have referred to some work presenting further details on inferential methods within our framework, for example on selection, process control and condition monitoring, and these appear to provide convincing opportunities for applications with attractive properties. However, further research is required to extent the area of application. For example, generalization to multivariate inference is crucial. Hill [27] suggests an approach which provides an interesting starting point for further analysis. Concepts from classical nonparametric statistics, for example 'statistically equivalent blocks' [1], might also provide ways towards such a generalization.

For most of this paper, predictive inferences were stated in terms of one future observation. It is of interest to consider multiple future observations, where also study of limit behaviour might provide new insights into the relation of our approach with, for example, likelihood methods. Such limits have been considered for Bernoulli data, from a similar perspective as presented in this paper [11]. As the random quantities representing the multiple future observations are assumed to be exchangeable, limits are related to De Finetti's representation theorem [21], providing different insights into the role of parameters occurring in statistical models when infinite exchangeability is assumed.

Finally, as we set out to study these $A_{(n)}$-based inferences, we only considered equal probabilities for the intervals of the partition of $\mathbb{R}$ as created by data $x_1, \ldots, x_n$. Our main results, however, can easily be generalized for different probabilities per interval. From the prospective of interval probability, it is interesting to replace the exact probabilities $1/(n+1)$ by interval probability, which might lead to inferences which also have attractive properties from a robust statistics perspective.

References

[1] Anderson, T.W. (1966). Some nonparametric procedures based on statistically equivalent blocks. In: P.R. Krishnaiah (ed.), *Multivariate Analysis I.* Academic Press, New York, 5-27.

[2] Arts, G.R.J., Coolen, F.P.A. and van der Laan, P. (2001). Nonparametric predictive inference in statistical process control. *Journal of Quality Technology*, to appear.

[3] Augustin, T. (1998). *Optimale Tests bei Intervallwahrscheinlichkeit.* Vandenhoeck & Ruprecht, Göttingen. (In German, with an English summary on pages 247-249.)

[4] Augustin, T. (2002). Neyman-Pearson testing under interval probability by globally least favorable pairs – reviewing Huber-Strassen theory and extending it to general interval probability. *Journal of Statistical Planning and Inference*, to appear.

[5] Berger, J.O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, **25**, 303-328.

[6] Boole, G. (1854). *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities.* Walton & Maberley, London. (Reprinted: Dover, New York, 1951.)

[7] Bernard, J.M. (ed.) (2002). Special issue on imprecise probabilities and their applications. *Journal of Statistical Planning and Inference*, to appear.

[8] Buja, A. (1986). On the Huber-Strassen Theorem. *Probability Theory and Related Fields*, **73**, 149-152.

[9] Chateauneuf, A., Cohen, M. and Meilijson, I. (1997). New tools to better model behavior under risk and uncertainty: an overview. *Finance*, **18**, 25-46.

[10] Choquet, G. (1954). Theory of capacities. *Annales de l'Institut Fourier*, **5**, 131-295.

[11] Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters*, **36**, 349-357.

[12] Coolen, F.P.A. and Coolen-Schrijner, P. (2000). Condition monitoring: a new perspective. *Journal of the Operational Research Society*, **51**, 311-319.

[13] Coolen, F.P.A. and van der Laan, P. (2001). Imprecise predictive selection based on low structure assumptions. *Journal of Statistical Planning and Inference*, **98**, 259-277.

[14] Cozman, F. and Moral, S. (eds.) (2000). Special volume on imprecise probabilities. *International Journal of Approximate Reasoning*, **24**, 121-299.

[15] Dempster, A.P. (1963). On direct probabilities. *Journal of the Royal Statistical Society B*, **25**, 100-110.

[16] Dempster, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, **37**, 325-339.

[17] de Cooman, G. (ed.) (2000). Special issue on imprecise probabilities. *Risk, Decision and Policy*, **5**, 107-181.

[18] de Cooman, G., Cozman, F.G., Moral, S. and Walley, P. (eds.) (1999). *ISIPTA'99: Proceedings of the First International Symposium on Imprecise Probabilities and their Applications.* University of Ghent.
`http://ippserv.rug.ac.be/~isipta99/index.html`

[19] de Cooman, G., Fine, T.L., Moral, S. and Seidenfeld, T. (eds.) (2001). *ISIPTA'01: Proceedings of the Second International Symposium on Imprecise Probabilities and*

their Applications. Cornell University, Ithaca, New York.
`http://decsai.ugr.es/~smc/isipta01/proceedings/index.html`

[20] de Cooman, G. and Walley, P. (eds.) (2001). *The Imprecise Probability Project.* `http://ippserv.rug.ac.be/`

[21] De Finetti, B. (1974). *Theory of Probability* (2 volumes). Wiley, London.

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

[23] Geisser, S. (1993). *Predictive Inference: an Introduction.* Chapman and Hall, London.

[24] Hampel, F. (1997). What can the foundations discussion contribute to data analysis? And what may be some of the future directions in robust methods and data analysis? *Journal of Statistical Planning and Inference*, **57**, 7-19.

[25] Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, **63**, 677-691.

[26] Hill, B.M. (1988). De Finetti's Theorem, Induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference (with discussion). In: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds.), *Bayesian Statistics 3.* University Press, Oxford, 211-241.

[27] Hill, B.M. (1993). Parametric models for A_n: splitting processes and mixtures. *Journal of the Royal Statistical Society B*, **55**, 423-433.

[28] Huber, P.J. (1981). *Robust Statistics.* Wiley, New York.

[29] Huber, P.J. and Strassen, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Annals of Statistics*, **1**, 251-263. (Correction: **2**, 223-224.)

[30] Lane, D.A. and Sudderth, W.D. (1984). Coherent predictive inference. *Sankhyā: The Indian Journal of Statistics, Series A*, **46**, 166-185.

[31] Papamarcou, A. and Fine, T.L. (1991). Unstable collectives and envelopes of probability measures. *The Annals of Probability*, **19**, 893-906.

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

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

[34] Strassen, V. (1964). Meßfehler und Information. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **2**, 273-305.

[35] Walley, P. (1981). *Coherent Lower (and Upper) Probabilities.* Technical Report, Department of Statistics, University of Warwick.

[36] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London.

[37] Walley, P. and Fine, T.L. (1982). Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, **10**, 741-761.

[38] Weichselberger, K. (1995). Axiomatic foundations of the theory of interval-probability. In: V. Mammitzsch and H. Schneeweiß (eds.), *Proceedings of the Second Gauß Symposion, Section B.* De Gruyter, Berlin, 47-64.

[39] Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, **24**, 149-170.

[40] Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept.* Physika, Heidelberg.

[41] Yager, R.R., Fedrizzi, M. and Kacprzyk, J. (eds.) (1994). *Advances in the Dempster-Shafer Theory of Evidence.* Wiley, New York.

[42] Yu, C. and Arasta, F. (1994). On conditional belief functions. *International Journal of Approximate Reasoning*, **10**, 155-172.