Bi-elastic Neighbourhood Models

A. WALLNER
Ludwig-Maximilians-University Munich, Germany

Abstract

We extend Buja's concept of "pseudo-capacities", which comprises the neighbourhood models for classical probabilities commonly used in robust statistics. Although systematically developing various directions for generalizing that model, we especially show that robust statistics can be freed from the severe restriction to 2-monotone capacities by employing the more natural framework of coherent or F-probabilities. Our main new tool for doing this is to use bi-elastic instead of convex functions.

Keywords

interval probability, robust statistics, neighbourhood models, distorted probability, pseudo-capacity, convex and bi-elastic functions

1 Introduction

The major concept in robust statistics for "robustifying" statements concerning classical distributions is to construct *neighbourhoods* of precise probabilities, which are called *central distributions* in this context. There is a famous method, due to Buja, accommodating, up to now, many of the corresponding neighbourhood models: Let p be some fixed classical probability, let $f\colon [0;1] \to [0;1]$ be a function with $f(0) = 0$ and $f(1) = 1$, and define

$$L = f \circ p. \tag{1}$$

By Denneberg (see [4], p. 17), a set function L constructed like this, is called a *distorted probability*, if f is increasing. In case $f(x) \le x$, $\forall x \in [0;1]$, L can be seen as the *lower bound* of an *interval probability*, which creates a neighbourhood of p in the sense that $L(A) \le p(A) \le U(A) := 1 - L(\neg A)$ for all events A.

Now in robust statistics the standard requirement concerning f is to be indeed *convex*. We suspect that nobody knows a reasonable philosophical argument, why this strong assumption is made. Instead it seems to have mere mathematical origins: "Only if f is convex, then L becomes an algebraic pushover." We want to convince the reader that not even this technical argument is true. Strictly speaking, the word "Only" should be replaced by "Not only".

If f is convex, then by Buja (cf. [3]) a set function L constructed in accordance with (1) is called a *pseudo-capacity*.[1] Now every pseudo-capacity is a *2-monotone* set function (see Theorem 1, model 5, and also [4], p. 17), and this fact seems to be the technical advantage. But from a philosophical point of view there are no visible reasons to restrict the frameworks of interval probability as well as of robust statistics to 2-monotonicity. Instead it is more natural to consider the wider class of Walley's *coherent probabilities* (cf. [8]), which are closely related to *F-probabilities* in the sense of Weichselberger (see [10] or [11]).

We will show that the formulation (1) is also useable for constructing the lower bound L of an F-probability, which is not necessarily 2-monotone. For this we have to weaken the condition of convexity for f and replace it by a new assumption: *bi-elasticity*.

Just as there exist 2-monotone set functions L, which cannot be described by (1) using convex functions f, we, of course, are not able to produce the whole class of F-probabilities by only employing the definition (1), letting p vary over all classical probabilities and f vary over all bi-elastic functions. But we will explain that bi-elasticity is exactly the *appropriate* requirement *when* defining F-probabilities via (1) (see Section 6). Moreover, from an algebraical point of view the generated subclass of F-probabilities is as easy manageable as the corresponding subclass of 2-monotone set functions, i.e. the class of pseudo-capacities.

In Section 2 we introduce the notion of bi-elasticity. In Section 3 a language for interval probability is fixed: As far as needed, we outline Weichselberger's formal and methodological framework. But this should be no restriction. Since, in particular, σ-additivity (instead of additivity) of classical probabilities does not play any role, the concepts developed could also be applied to other theories of imprecise probabilities, especially to Walley's theory. In Section 4 we go into the details of the convex and bi-elastic neighbourhood models described above, resulting in Theorem 1. There we, in fact, will not use the phrasing of equation (1): Since sometimes it is necessary to apply methods of robustness to interval probability itself[2] and, anyway, it is a natural mathematical task to look for closure properties, we consider the more generalized form

$$L = f \circ L_0, \tag{2}$$

where L_0 is the lower bound of some given *interval-valued central distribution*. Learning from Theorem 1, we also deal with a modified version of it, which is stated in Theorem 2. Its formulation serves, in essence, as a motivation for Section 5, i.e. for Theorems 3 and 4, which significantly generalize the neighbourhood models developed before. Section 6 is reserved for concluding remarks.

To give reasons for the successive steps, the structure of this technical paper is rather heuristic. Hence the proofs are postponed repeatedly — until the proof of the last theorem.

[1] See [2] for more detailed information.

[2] See [1], pp. 229ff, for a discussion of this topic.

2 Convex and Bi-elastic Functions

What is *bi-elasticity*? Suppose we concentrate on a function f: $[0;1] \to [0;1]$ with $f(0) = 0$ and $f(1) = 1$ and imagine that three points of f's graph, namely $(x, f(x))$, $(y, f(y))$, and $(z, f(z))$ with $0 \le x < y < z \le 1$, are *standing in convex position*. Obviously, by this terminology we mean that the following local comparison of quotients of differences is valid:

$$\frac{f(y) - f(x)}{y - x} \le \frac{f(z) - f(y)}{z - y}. \tag{3}$$

If this inequality is globally true, i.e. *for all* such x, y, z, we usually say that f is *convex*. Now fix $x = 0$, and let just y and z vary. Then it is easily seen that we get equivalently

$$\frac{f(y)}{y} \le \frac{f(z)}{z}, \quad \forall y, z \text{ with } 0 < y \le z \le 1, \tag{4}$$

i.e. that the *average of f is increasing*. In economic sciences this behaviour of f is called *elastic* (e.g. see [5]).

So, what's *bi*-elasticity? For this new concept (introduced in [9], Chapter 6), let first f be elastic, and secondly set $z = 1$ in (3) and let x and y vary. After simple transformations we get

$$\frac{1 - f(x)}{1 - x} \le \frac{1 - f(y)}{1 - y}, \quad \forall x, y \text{ with } 0 \le x \le y < 1, \tag{5}$$

as an equivalent form, which, in turn, is equivalent to

$$\frac{1 - f(1-y)}{y} \le \frac{1 - f(1-x)}{x}, \quad \forall x, y \text{ with } 0 < x \le y \le 1. \tag{6}$$

Thus, additionally, the *conjugate function of f*, i.e. $x \mapsto 1 - f(1-x)$, has to have *decreasing average*. We summarize:

Definition 1 Let f: $[0;1] \to [0;1]$ with $f(0) = 0$ and $f(1) = 1$. Then f is called

1. *convex*, if (3) holds for all x, y, z with $0 \le x < y < z \le 1$,
2. *bi-elastic*, if (4) and (6) are valid. □

Corollary 1 *Let f: $[0;1] \to [0;1]$ with $f(0) = 0$ and $f(1) = 1$.*

1. *f is convex iff $f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y)$, $\forall x, y, \lambda \in [0;1]$.*
2. *f is bi-elastic iff $f(\lambda x) \le \lambda f(x)$ and $\lambda(1 - f(1-x)) \le 1 - f(1-\lambda x)$, $\forall x, \lambda \in [0;1]$.*
3. *If f is convex, then f is bi-elastic.*
4. *If f is bi-elastic, then $f(x) \le x$, $\forall x \in [0;1]$.*[3] □

Proof. 1.) and 2.) can be shown straightforwardly. For 3.) see above, for 4.) put $y = x$ and $z = 1$ in (4). □

[3] Moreover, every bi-elastic function is monotone in $[0;1]$ and continuous in $[0;1[$.

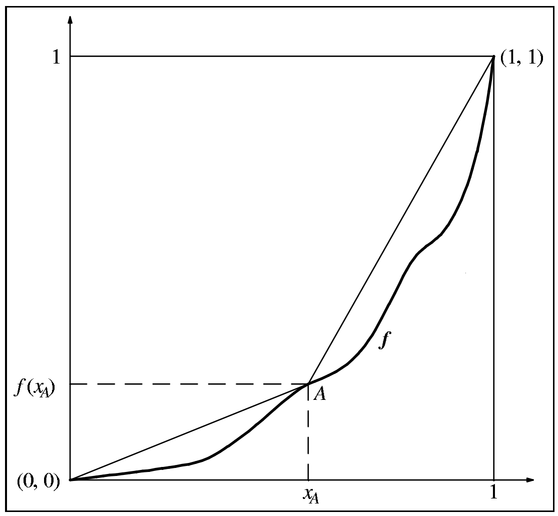

Figure 1: An example of a bi-elastic function f. Bi-elasticity of f can be described equivalently as follows: For each point $A = (x_A, f(x_A))$ on the graph of f, the graph of f between 0 and x_A nowhere is lying above the line between $(0,0)$ and A, and between x_A and 1 it nowhere is lying above the line between A and $(1, 1)$.

3 Basic Definitions of Interval Probability according to Weichselberger

Here we report the main concepts of Weichselberger's theory of interval probability (see [10] or [11]), adding some slight modifications. For the following let Ω be a fixed *sample space* and $\mathcal{A}$ a fixed σ-*algebra* over Ω. Hence $(\Omega; \mathcal{A})$ is fixed *measurable space*.

Definition 2 A set function $p\colon \mathcal{A} \to [0; 1]$ is called a *K-function* (classical probability) on $(\Omega; \mathcal{A})$, if it satisfies the axioms of Kolmogorov. The set of all K-functions on $(\Omega; \mathcal{A})$ is denoted by $\mathcal{K}(\Omega; \mathcal{A})$. □

Definition 3

1. A triple $O = (\Omega; \mathcal{A}; L)$ is called an *adjusted O-field*, if $L\colon \mathcal{A} \to [0; 1]$ is a set function, which is *normed*, i.e. $L(\emptyset) = 0$ and $L(\Omega) = 1$. The set $\mathcal{M}(O) = \{p \in \mathcal{K}(\Omega; \mathcal{A}) \mid L(A) \leq p(A),\ \forall A \in \mathcal{A}\}$ is called the *structure of O*.

2. An adjusted O-field $\mathcal{R}$ is called an *adjusted R-(probability) field*, if $\mathcal{M}(\mathcal{R}) \neq \emptyset$.

3. An adjusted R-field $\mathcal{F} = (\Omega; \mathcal{A}; L)$ is called an *F-(probability) field*, if it satisfies the axiom $L(A) = \inf_{p \in \mathcal{M}(\mathcal{F})} p(A), \forall A \in \mathcal{A}$.[4]
4. An adjusted R-field $\mathcal{F} = (\Omega; \mathcal{A}; L)$ is called an *F_0-(probability) field*, if it satisfies the axiom $L(A) = \min_{p \in \mathcal{M}(\mathcal{F})} p(A), \forall A \in \mathcal{A}$.
5. An adjusted O-field $(\Omega; \mathcal{A}; L)$ is called a *CA-field*, if L is *2-monotone*, i.e. $L(A) + L(B) \leq L(A \cup B) + L(A \cap B), \forall A, B \in \mathcal{A}$.
6. A CA-field is called a *C-(probability) field*, if it is an F-field.
7. A CA-field is called a *C_0-(probability) field*, if it is an F_0-field.
8. A triple $(\Omega; \mathcal{A}; p)$ is called a *K-(probability) field*, if p is a K-function. □

Since $(\Omega; \mathcal{A})$ is fixed, every adjusted O-field $O = (\Omega; \mathcal{A}; L)$ is determined by the “lower bound” L. Subsequently we always “associate” the “upper bound” U of O via *conjugation* of L, i.e. $U(.) = 1 - L(\neg.)$.

Some comments on Definition 3 are useful:

- Weichselberger’s original definition of an *R-field* is that of a quadruple $\mathcal{R} = (\Omega; \mathcal{A}; L, U)$ having a non-empty structure $\mathcal{M}(\mathcal{R}) = \{p \in \mathcal{K}(\Omega; \mathcal{A}) \mid L(A) \leq p(A) \leq U(A), \forall A \in \mathcal{A}\}$. In this setting neither L is normed necessarily, nor L and U have to be conjugate, what both is not appropriate for our purposes.
- In [1], Corollary 2.13, it is shown that every *continuous* F-field is an F_0-field. (Hence, in particular, every F-field on a finite measurable space has the F_0-property.) Since, on the one hand, we don’t want to discuss topological features here, but, on the other hand, intend to deal with closure properties concerning F-fields as well as F_0-fields, we distinguish both cases by introducing these two terms.
- It is known that every CA-field is a C_0-field, and hence a C-field, in case the sample space Ω is finite. For the general case, usually additional topological assumptions are made to enforce the F-(or F_0-)property, in particular, for defining *2-monotone capacities* (cf. [7]). But, as mentioned above, we want to abstain from topological aspects here. So the CA-property, i.e., essentially, the 2-monotonicity of the lower bound, should be considered as the extracted pure algebraic part of the definition of C-(or C_0-)fields. There are some closure properties, we want to emphasize later, only concerning this algebraic part. Therefore the definitions of CA-, C-, and C_0-fields are organized as stated.

[4] The definitions of adjusted R-fields and of F-fields are closely related to Walley’s *avoiding sure loss* and *coherence* respectively (cf. [8]).

For later use we record the following corollary, which can be proven straightforwardly.

Corollary 2

1. *If $O = (\Omega; \mathcal{A}; L)$ is an adjusted O-field and $U(.) = 1 - L(\neg .)$, then $\mathcal{M}(O) = \{p \in \mathcal{K}(\Omega; \mathcal{A}) \mid p(A) \leq U(A),\ \forall A \in \mathcal{A}\}$.*

2. *If $(\Omega; \mathcal{A}; L)$ is an F- or a CA-field, then L and its conjugate U are* monotone*, i.e., for $\Psi \in \{L, U\}$ we have $\forall A, B \in \mathcal{A}$: $A \subseteq B \Longrightarrow \Psi(A) \leq \Psi(B)$.*

3. *If $O_1 = (\Omega; \mathcal{A}; L_1)$ and $O_2 = (\Omega; \mathcal{A}; L_2)$ are adjusted O-fields, then*

$$L_1(.) \leq L_2(.) \Longrightarrow \mathcal{M}(O_2) \subseteq \mathcal{M}(O_1). \qquad \square$$

As a mnemonic device concerning the definitions above, we get the following clear picture:

$$\begin{array}{ccccccc}
 & & \text{CA-f.} \wedge \text{F}_0\text{-f.} & \Rightarrow & \text{F}_0\text{-field} & & \\
 & & \Updownarrow & & \Downarrow & & \\
\text{K-field} & \Rightarrow & \text{C}_0\text{-field} & & \text{F-field} & \Rightarrow \text{adj. R-field} \Rightarrow & \text{adj. O-field.} \\
 & & \Downarrow & & \Uparrow & & \\
 & & \text{C-field} & \Leftrightarrow & \text{CA-f.} \wedge \text{F-f.} & &
\end{array}$$

4 Convex and Bi-elastic Neighbourhood Models

For constructing *neighbourhoods* of classical probabilities, in robust statistics mainly metrics are used to define appropriate topologies over the space $\mathcal{K}(\Omega;\mathcal{A})$ (e.g. see [6]). Here we do not rely on the term "neighbourhood" in some topological sense, and that is why we give the trivial

Definition 4 For adjusted O-fields $O_0 = (\Omega; \mathcal{A}; L_0)$, $O = (\Omega; \mathcal{A}; L)$ and a K-function p, we say that

- O is a *neighbourhood of* O_0, if $L(.) \leq L_0(.)$,
- O is a *neighbourhood of* p, if O is a neighbourhood of $(\Omega; \mathcal{A}; p)$. $\square$

Therefore, O is a neighbourhood of the K-function p iff simply p is an element of the structure of O, and hence O is an adjusted R-field. In general, we have $\mathcal{M}(O_0) \subseteq \mathcal{M}(O)$, if O is a neighbourhood of O_0 (cf. Corollary 2, 3.)).

Now we come to a first category of neighbourhood models motivated in Section 1. Inspired by the notions of *pseudo-capacities* (the starting point of our developments), *bi-elastic functions* (generalizing convex functions), and *interval-valued central distributions* (including precise central distributions as a specific case), we get

Theorem 1 ***(First class of neighbourhood models)*** *Let $L_0\colon \mathcal{A} \to [0;1]$ be a set function, $f\colon [0;1] \to [0;1]$ a function with $f(0) = 0$ and $f(1) = 1$, $L = f \circ L_0$, $\mathcal{O}_0 = (\Omega; \mathcal{A}; L_0)$, and $\mathcal{O} = (\Omega; \mathcal{A}; L)$. Then we have:*[5]

1. *If $\mathcal{O}_0$ is an adjusted O-field, then so is $\mathcal{O}$.*
2. *If $f(x) \le x$, $\forall x \in [0;1]$, and $\mathcal{O}_0$ is an adjusted R-field, then so is $\mathcal{O}$.*
3. *If f is bi-elastic, and $\mathcal{O}_0$ is an F-field, then so is $\mathcal{O}$.*
4. *If f is bi-elastic, and $\mathcal{O}_0$ is an F_0-field, then so is $\mathcal{O}$.*
5. *If f is convex, and $\mathcal{O}_0$ is a CA-field, then so is $\mathcal{O}$.*
6. *If f is convex, and $\mathcal{O}_0$ is a C-field, then so is $\mathcal{O}$.*
7. *If f is convex, and $\mathcal{O}_0$ is a C_0-field, then so is $\mathcal{O}$.*

Moreover, in the cases 2.)–7.) $\mathcal{O}$ is a neighbourhood of $\mathcal{O}_0$. □

Proof. 1.) and 2.) are obvious. For 3.)–7.) see Theorem 2 below.[6] The "Moreover"-statement follows from Corollary 1, 3.) and 4.). □

From now on we concentrate on the most interesting cases, namely F-, F_0-, CA-, C-, and C_0-fields. Our goal is to generalize models 3–7 of Theorem 1 in two steps, which leads to Theorems 2, 3, and 4.

The *first step* is just a small one and is based on an elementary observation. Let us for the moment consider model 5 of Theorem 1: In order to maintain the 2-monotonicity, we, in essence, made two assumptions: the definition of L, i.e. $L = f \circ L_0$, and the convexity of f. By Definition 1, 1.), this implies

$$\frac{L(B) - L(A)}{L_0(B) - L_0(A)} \le \frac{L(C) - L(B)}{L_0(C) - L_0(B)}, \tag{7}$$

for all A, B, $C \in \mathcal{A}$ with $L_0(A) < L_0(B) < L_0(C)$. Now it is natural to suspect that it doesn't matter, how f is defined on $[0;1] \setminus \{L_0(A) \mid A \in \mathcal{A}\}$. It should be sufficient for our CA-model to presuppose the inequalities (7). Similarly, we expect that models 3 and 4 of Theorem 1 could be modified analogously: The corresponding inequalities given by bi-elasticity are (cf. (4) and (5))

$$\frac{L(A)}{L_0(A)} = \frac{f(L_0(A))}{L_0(A)} \le \frac{f(L_0(B))}{L_0(B)} = \frac{L(B)}{L_0(B)},$$

for all A, $B \in \mathcal{A}$ with $0 < L_0(A) \le L_0(B)$, and, additionally, using $U_0(.) = 1 - L_0(\neg .)$ and $U(.) = 1 - L(\neg .)$,

$$\frac{U(B)}{U_0(B)} = \frac{1-L(\neg B)}{1-L_0(\neg B)} = \frac{1-f(L_0(\neg B))}{1-L_0(\neg B)} \le \frac{1-f(L_0(\neg A))}{1-L_0(\neg A)} = \frac{1-L(\neg A)}{1-L_0(\neg A)} = \frac{U(A)}{U_0(A)},$$

for all $A, B \in \mathcal{A}$ with $L_0(\neg B) \le L_0(\neg A) < 1$, i.e., equivalently, $0 < U_0(A) \le U_0(B)$.

These considerations are summed up in

[5] Models 5–7 reflect the concept of pseudo-capacities, in case of a precise central distribution $\mathcal{O}_0$.

[6] For the moment, we can say that 6.) is a consequence of 3.) and 5.), and 7.) is a consequence of 4.) and 5.), since convexity implies bi-elasticity (cf. Corollary 1, 3.)).

Theorem 2 ***(Second class of neighbourhood models)*** *Let $O_0 = (\Omega;\mathcal{A};L_0)$ and $O = (\Omega;\mathcal{A};L)$ be adjusted O-fields, $U_0(.) = 1-L_0(\neg.)$, and $U(.) = 1-L(\neg.)$.*

1. *Suppose that O_0 is an F-field and that the following two conditions hold:*

   (a) $$L(A)\cdot L_0(B) \le L_0(A)\cdot L(B), \quad \forall A,B\in\mathcal{A} \text{ with } L_0(A)\le L_0(B); \tag{8}$$

   (b) $$U_0(A)\cdot U(B) \le U(A)\cdot U_0(B), \quad \forall A,B\in\mathcal{A} \text{ with } U_0(A)\le U_0(B). \tag{9}$$

   Then O is an F-field, too.

2. *Suppose that O_0 is an F_0-field and that conditions (8) and (9) hold. Then O is an F_0-field, too.*

3. *Suppose that O_0 is a CA-field and that the following condition holds:*[7]

   $$(L(B)-L(A))\cdot(L_0(C)-L_0(B)) \le (L_0(B)-L_0(A))\cdot(L(C)-L(B)), \quad \forall A,B,C\in\mathcal{A} \text{ with } L_0(A)\le L_0(B)\le L_0(C). \tag{10}$$

   Then O is a CA-field, too.

4. *Suppose that O_0 is a C-field and that condition (10) holds. Then O is a C-field, too.*

5. *Suppose that O_0 is a C_0-field and that condition (10) holds. Then O is a C_0-field, too.*

Moreover, in all five cases we have: O is a neighbourhood of O_0, and the "functional connection"

$$\forall A,B\in\mathcal{A}: \quad L_0(A) = L_0(B) \Longrightarrow L(A) = L(B) \tag{11}$$

between L_0 and L is valid. □

Proof. It is straightforward that from condition (10) we can derive conditions (8) and (9) (for (8) put $A=\emptyset$ in (10), for (9) set $C=\Omega$ in (10)[8]). Hence, on the one hand, 4.) is a direct consequence of 1.) and 3.), and 5.) is a consequence of 2.) and 3.). On the other hand, the "Moreover"-statement can be deduced from (8): By putting $B=\Omega$, we get

$$L(A)\le L_0(A), \quad \forall A\in\mathcal{A}, \tag{12}$$

which is the statement that O is a neighbourhood of O_0. To prove (11), let $L_0(A) = L_0(B)$. By (12), we can assume $L_0(B) > 0$. But then, two applications of (8) lead to $L(A)\cdot L_0(B) = L_0(A)\cdot L(B) = L_0(B)\cdot L(B)$, thus $L(A) = L(B)$.

Summarizing, we have shown all parts of Theorem 2 — with the exception of its heart: statements 1.), 2.), and 3.). For this we refer to Theorem 3 below, since in the situations of 1.), 2.), and 3.) the set functions L_0 and U_0 are monotone (cf. Corollary 2, 2.)). To be complete, we have to prove the additional premise (15) in Theorem 3. But this is an easy result of (11) (just set $B=\Omega$), which is proved already. □

[7] In (10) it's not sufficient to use quotients as above, excluding the possibility that the denominator is 0.

[8] We can argue in a manner similar to the proof sketch, given at the beginning of Section 2, where we deduced bi-elasticity from convexity.

Clearly, the most important case of Theorem 2 is that the central distribution $\mathcal{O}_0$ is some K-field $(\Omega; \mathcal{A}; p_0)$. For this we give an example, which — historically — led to all generalized neighbourhood models presented here.

Example 1 Let $(\Omega; \mathcal{A}) = (\Omega_k; \mathcal{P}(\Omega_k))$ be a finite measurable space, where $|\Omega_k| = k \in \mathbb{N}$ and $\mathcal{P}(\Omega_k)$ is the power set of Ω_k. We consider the consequences of Theorem 2 for the case $\mathcal{O}_0 = (\Omega_k; \mathcal{P}(\Omega_k); p_0^k)$, in which p_0^k is the *classical uniform probability* on $(\Omega_k; \mathcal{P}(\Omega_k))$, i.e. $p_0^k(A) = \frac{|A|}{k}$, $\forall A \subseteq \Omega_k$. Let $\mathcal{O} = (\Omega_k; \mathcal{P}(\Omega_k); L)$ be some adjusted O-field. From (11) we conclude

$$\forall A, B \subseteq \Omega_k : \quad |A| = |B| \Longrightarrow L(A) = L(B), \tag{13}$$

which means that the only possibility in generating $\mathcal{O}$ as a neighbourhood of p_0^k with the methods of Theorem 2, we have to restrict ourselves to *uniform interval probability*. Hence we assume (13) and write for $i = 0, \ldots, k$: $L^{(i)} = L(A)$, if $i = |A|$ for some $A \subseteq \Omega_k$, and consistently $U^{(i)} = 1 - L^{(k-i)}$. Additionally, we concentrate on considering models 1 and 2 of Theorem 2, the F- and the F_0-model, which are the same, since Ω_k is finite. Conditions (8) and (9) are equivalent to the chain

$$\frac{L^{(1)}}{1} \le \frac{L^{(2)}}{2} \le \cdots \le \frac{L^{(k-1)}}{k-1} \le \frac{1}{k} \le \frac{U^{(k-1)}}{k-1} \le \cdots \le \frac{U^{(2)}}{2} \le \frac{U^{(1)}}{1}. \tag{14}$$

Therefore, model 1 of Theorem 2 says: Every adjusted uniform O-field $\mathcal{O} = (\Omega_k; \mathcal{P}(\Omega_k); L)$ is an F-field — an "*uniform F-field*" —, if it obeys the chain (14). In [11], Lemma 4.3.5, it is shown over and above that, that (14) is also necessary for $\mathcal{O}$ to be an uniform F-field on $(\Omega_k; \mathcal{P}(\Omega_k))$. □

Theorem 2 is only a very slight generalization of the models 3–7 in Theorem 1. For example, if condition (10) holds, it always is possible to construct a convex function f: $[0; 1] \to [0; 1]$ with $f(0) = 0$ and $f(1) = 1$ such that $L = f \circ L_0$. Similarly for (8) and (9) on the one hand and bi-elastic functions defined on $[0; 1]$ on the other hand.

Theorem 2 should rather be seen as a motivation for Theorem 3 given in the next section.

5 Generalized Convex and Bi-elastic Neighbourhood Models

The inequalities, working as premises in conditions (8), (9), and (10) do not seem to be very natural. For example, in (8) it would be nice to replace "$L_0(A) \le L_0(B)$" by "$A \subseteq B$", since then, e.g., we would have a connection to conditional interval probability (see Section 6).

Let us formulate this big *second step* of generalizing the neighbourhood models, fundamentally first presented in [9], Chapter 6:

Theorem 3 ***(Third class of neighbourhood models, part 1)*** *Let $O_0 = (\Omega;\mathcal{A};L_0)$ and $O = (\Omega;\mathcal{A};L)$ be adjusted O-fields, $U_0(.) = 1-L_0(\neg.)$, and $U(.) = 1-L(\neg.)$. Assume additionally that we have*[9]

$$\forall A \in \mathcal{A}: \quad L_0(A) = 1 \Longrightarrow L(A) = 1. \tag{15}$$

1. *Suppose that O_0 is an F-field and that the following two conditions hold:*

   (a) $L(A)\cdot L_0(B) \le L_0(A)\cdot L(B), \quad \forall A,B \in \mathcal{A}$ *with* $A \subseteq B$; (16)

   (b) $U_0(A)\cdot U(B) \le U(A)\cdot U_0(B), \quad \forall A,B \in \mathcal{A}$ *with* $A \subseteq B$. (17)

   Then O is an F-field, too.

2. *Suppose that O_0 is an F_0-field and that conditions (16) and (17) hold. Then O is an F_0-field, too.*

3. *Suppose that O_0 is a CA-field and that the following condition holds:*

$$(L(B)-L(A))\cdot(L_0(C)-L_0(B)) \le (L_0(B)-L_0(A))\cdot(L(C)-L(B)), \\ \forall A,B,C \in \mathcal{A} \text{ with } A \subseteq B \subseteq C. \tag{18}$$

   Then O is a CA-field, too.

4. *Suppose that O_0 is a C-field and that condition (18) holds. Then O is a C-field, too.*

5. *Suppose that O_0 is a C_0-field and that condition (18) holds. Then O is a C_0-field, too.*

Moreover, in all five cases we have: O is a neighbourhood of O_0, and the "functional connection"

$$\forall A,B \in \mathcal{A}: \quad A \subseteq B \wedge L_0(A) = L_0(B) \Longrightarrow L(A) = L(B) \tag{19}$$

between L_0 and L is valid. □

Proof. Let O_0 and O be adjusted O-fields as denoted. First we prove:

1. (16) $\Longrightarrow$ $L(A) \le L_0(A)$, $\forall A \in \mathcal{A}$.
2. (16) $\Longrightarrow$ $\big(\forall A,B \in \mathcal{A}: A \subseteq B \wedge L_0(A) \le L_0(B) \Longrightarrow L(A) \le L(B)\big)$.
3. (15) $\wedge$ (16) $\wedge$ (17) $\Longrightarrow$ (19).

[9] It can easily be seen that the models don't work, if we drop this additional condition. (15) is equivalent with $\forall A \in \mathcal{A}: U_0(A) = 0 \Rightarrow U(A) = 0$, and hence with

$$\forall A \in \mathcal{A}: \ (\forall p_0 \in \mathcal{M}(O_0).\ p_0(A) = 0) \Longrightarrow (\forall p \in \mathcal{M}(O).\ p(A) = 0),$$

which means that O is *absolutely continuous with respect to O_0*.

For a), just let $B=\Omega$ in (16). For b) assume $A\subseteq B$ and $L_0(A)\le L_0(B)$, where by a) w.l.o.g. $L_0(A)>0$. Together with (16) we get $L(A)\cdot L_0(A)\le L(A)\cdot L_0(B)\le L_0(A)\cdot L(B)$, hence $L(A)\le L(B)$. For c) suppose (15), (16), (17), $A\subseteq B$, and $L_0(A)=L_0(B)$, thus also $\neg B\subseteq\neg A$ and $U_0(\neg A)=U_0(\neg B)$. By (15), w.l.o.g. we can assume $L_0(B)<1$, hence $U_0(\neg B)>0$. Now (17) gives $U_0(\neg B)\cdot U(\neg A)\le U(\neg B)\cdot U_0(\neg A)=U(\neg B)\cdot U_0(\neg B)$, thus $U(\neg A)\le U(\neg B)$, i.e. $L(B)\le L(A)$. Together with b) we infer $L(A)=L(B)$.

Now we give the proof of Theorem 3. First it can easily be seen that (18) implies (16) and (17) (let $A=\emptyset$ or $C=\Omega$ in (18)). Therefore, on the one hand, the "Moreover"-statement is a trivial conclusion of a) and c), and, on the other hand, 4.) is a consequence of 1.) and 3.), and 5.) is a consequence of 2.) and 3.).

For 1.) and 2.) we refer to Theorem 4 (see below).

So here we just have to prove 3.), i.e., we have to show that condition (18) transfers 2-monotonicity from L_0 to L. For this, let (18) be valid and L_0 be 2-monotone. Then, according to Corollary 2, 2.), L_0 is monotone 2, and by b) we also infer the monotonicity of L. Now let $A,B\in\mathcal{A}$ be given. We have to show that

$$L(A)+L(B)\le L(A\cup B)+L(A\cap B). \tag{20}$$

If $L_0(A\cap B)=L_0(A)$, then by (19) $L(A\cap B)=L(A)$, hence (20) follows from the monotonicity of L. Thus we assume $L_0(A\cap B)<L_0(A)$ and, symmetrically, $L_0(A\cap B)<L_0(B)$. But then, by the 2-monotonicity of L_0 we have $L_0(A)<L_0(A\cup B)$ and $L_0(B)<L_0(A\cup B)$. Together with (18), we infer for $X\in\{A,B\}$:

$$0\le x(X):=\frac{L(X)-L(A\cap B)}{L_0(X)-L_0(A\cap B)}\le\frac{L(A\cup B)-L(X)}{L_0(A\cup B)-L_0(X)}=:y(X).$$

Now, by symmetric reasons, we suppose that $y(A)\le y(B)$, hence $x(A)\le y(B)$. Finally, the 2-monotonicity of L_0 leads to $L(A)-L(A\cap B)=x(A)\cdot(L_0(A)-L_0(A\cap B))\le x(A)\cdot(L_0(A\cup B)-L_0(B))\le y(B)\cdot(L_0(A\cup B)-L_0(B))=L(A\cup B)-L(B)$, thus (20) holds. □

The proof of Theorem 3 is not complete, because models 1 and 2 are waiting for verification. The reason for this is that we want to emphasize that these models are, in fact, *local* models with respect to the F- and F_0-property respectively.[10] This is the content of the following Theorem 4, the last one in the sequence of theorems.

Definition 5 Let $A\in\mathcal{A}$ be fixed. An adjusted R-field $\mathcal{R}=(\Omega;\mathcal{A};L)$ is called

1. an *F(A)-field*, if it satisfies the axiom $L(A)=\inf_{p\in\mathcal{M}(\mathcal{R})}p(A)$,
2. an *$F_0(A)$-field*, if it satisfies the axiom $L(A)=\min_{p\in\mathcal{M}(\mathcal{R})}p(A)$. □

The trivial connection with Definition 3, 3.) and 4.), is given by

Corollary 3 *Let $\mathcal{R}=(\Omega;\mathcal{A};L)$ be an adjusted R-field. Then we have:*

1. *$\mathcal{R}$ is an F-field iff for all $A\in\mathcal{A}$, $\mathcal{R}$ is an F(A)-field.*
2. *$\mathcal{R}$ is an F_0-field iff for all $A\in\mathcal{A}$, $\mathcal{R}$ is an $F_0(A)$-field.* □

[10] This also is true for the F- and F_0-models in Theorems 1 and 2.

Theorem 4 ***(Third class of neighbourhood models, part 2)*** *Let $A \in \mathcal{A}$ be fixed. Let $O_0 = (\Omega; \mathcal{A}; L_0)$ and $O = (\Omega; \mathcal{A}; L)$ be adjusted O-fields, $U_0(.) = 1 - L_0(\neg.)$, and $U(.) = 1 - L(\neg.)$. Assume that (15), (16), and (17) hold. Then we have:*

1. *If O_0 is an F(A)-field, then so is O.*
2. *If O_0 is an $F_0(A)$-field, then so is O.*

Moreover, in both cases O is a neighbourhood of O_0, and the "functional connection" (19) between L_0 and L is valid. □

Proof. The "Moreover"-statement can be shown like a) and c) in the proof of Theorem 3. So we only have to prove statements 1.) and 2.), where we restrict ourselves to model 1.[11] For this let all the corresponding premises be given, especially let $A \in \mathcal{A}$ be fixed and O_0 be an F(A)-field. Since O is a neighbourhood of O_0, we have

$$L(.) \leq L_0(.) \text{ and } U_0(.) \leq U(.), \quad \text{and thus} \quad \mathcal{M}(O_0) \subseteq \mathcal{M}(O). \tag{21}$$

(Hence the R-property moves from O_0 to O.) Now we concentrate on proving the F(A)-property of O, where by (21) w.l.o.g. we assume $L(A) < L_0(A)$. Together with (15) we infer

$$U(\neg A) > U_0(\neg A) > 0. \tag{22}$$

Let $\varepsilon > 0$, w.l.o.g.

$$\varepsilon < U(\neg A) - U_0(\neg A). \tag{23}$$

We have to show that there exists $p \in \mathcal{M}(O)$ with $p(A) \leq L(A) + \varepsilon$. Define

$$\delta = \frac{U_0(\neg A)}{U(\neg A)} \cdot \varepsilon. \tag{24}$$

Then, by (22), $\delta > 0$. Since O_0 is an F(A)-field, there is

$$p_0 \in \mathcal{M}(O_0) \quad \text{with} \quad p_0(A) \leq L_0(A) + \delta. \tag{25}$$

Together with (22), (23), and (24) we get by easy calculations

$$0 < U_0(\neg A) - \delta \leq p_0(\neg A) \leq U_0(\neg A) < U(\neg A) - \varepsilon \leq 1. \tag{26}$$

Therefore

$$1 \leq \frac{U(\neg A) - \varepsilon}{p_0(\neg A)} \leq \frac{U(\neg A) - \varepsilon}{U_0(\neg A) - \delta} = \frac{U(\neg A)}{U_0(\neg A)}, \tag{27}$$

where (24) is used for the equality. In addition, (26) implies that $p_0(A)$ and $p_0(\neg A)$ have positive values, and hence it is possible to define the classical conditional probabilities

$$p_0(. \mid A) = \frac{p_0(A \cap .)}{p_0(A)} \quad \text{and} \quad p_0(. \mid \neg A) = \frac{p_0(\neg A \cap .)}{p_0(\neg A)}.$$

Now we let

$$p(.) = (L(A) + \varepsilon) \cdot p_0(. \mid A) + (U(\neg A) - \varepsilon) \cdot p_0(. \mid \neg A), \tag{28}$$

which (using (26)) is a convex combination of $p_0(. \mid A)$ and $p_0(. \mid \neg A)$. Hence p is a well-defined K-function on $(\Omega; \mathcal{A})$. Moreover, we have $p(A) = L(A) + \varepsilon$.

To verify that p is an element of the structure of O, let $B \in \mathcal{A}$. We have to prove that $p(B) \geq L(B)$, where w.l.o.g. $L(B) > 0$. But then, by (21) we also have $L_0(B) > 0$. In addition, $L_0(A \cup B) > 0$.[12] From (16) we derive $\frac{L(A \cup B)}{L_0(A \cup B)} \geq \frac{L(B)}{L_0(B)}$, thus with (25),

[11] Modifying the following arguments by setting $\varepsilon = \delta = 0$, we also get a proof of model 2.

[12] Note that we are not able to infer this inequality from $L_0(B) > 0$ by monotonicity of L_0, since

$$p_0(B) \cdot \frac{L(A \cup B)}{L_0(A \cup B)} \geq L(B). \tag{29}$$

Furthermore, using the abbreviation

$$\Delta = U_0(\neg A \cap \neg B) - p_0(\neg A \cap \neg B) = p_0(A \cup B) - L_0(A \cup B), \tag{30}$$

we get the following inequalities, where (31) follows from (27) and (17), (32) is a consequence of (25), and (33) is implied by (27) and (21):

$$U_0(\neg A \cap \neg B) \cdot \frac{U(\neg A) - \varepsilon}{p_0(\neg A)} \leq U(\neg A \cap \neg B), \tag{31}$$

$$\Delta \geq 0, \tag{32}$$

$$\frac{U(\neg A) - \varepsilon}{p_0(\neg A)} \geq \frac{L(A \cup B)}{L_0(A \cup B)}. \tag{33}$$

If $\frac{L(A)+\varepsilon}{p_0(A)} > \frac{L(A \cup B)}{L_0(A \cup B)}$,[13] we calculate

$$\begin{aligned}
p(B) &\overset{(28)}{=} p_0(A \cap B) \cdot \frac{L(A)+\varepsilon}{p_0(A)} + p_0(\neg A \cap B) \cdot \frac{U(\neg A) - \varepsilon}{p_0(\neg A)} \\
&\overset{(33)}{\geq} p_0(A \cap B) \cdot \frac{L(A \cup B)}{L_0(A \cup B)} + p_0(\neg A \cap B) \cdot \frac{L(A \cup B)}{L_0(A \cup B)} \\
&= p_0(B) \cdot \frac{L(A \cup B)}{L_0(A \cup B)} \overset{(29)}{\geq} L(B).
\end{aligned}$$

Therefore, we can assume

$$\frac{L(A)+\varepsilon}{p_0(A)} \leq \frac{L(A \cup B)}{L_0(A \cup B)}. \tag{34}$$

Now we receive

$$\begin{aligned}
p(B) &= 1 - p(\neg B) \\
&\overset{(28)}{=} 1 - (L(A)+\varepsilon) \cdot \frac{p_0(A \cap \neg B)}{p_0(A)} - (U(\neg A) - \varepsilon) \cdot \frac{p_0(\neg A \cap \neg B)}{p_0(\neg A)} \\
&\overset{(30)}{=} 1 - U_0(\neg A \cap \neg B) \cdot \frac{U(\neg A) - \varepsilon}{p_0(\neg A)} + \Delta \cdot \frac{U(\neg A) - \varepsilon}{p_0(\neg A)} - p_0(A \cap \neg B) \cdot \frac{L(A)+\varepsilon}{p_0(A)} \\
&\overset{(31)\text{–}(34)}{\geq} 1 - U(\neg A \cap \neg B) + \Delta \cdot \frac{L(A \cup B)}{L_0(A \cup B)} - p_0(A \cap \neg B) \cdot \frac{L(A \cup B)}{L_0(A \cup B)} \\
&\overset{(30)}{=} p_0(B) \cdot \frac{L(A \cup B)}{L_0(A \cup B)} \overset{(29)}{\geq} L(B).
\end{aligned}$$

Hence Theorem 4 is proven. □

we did not presuppose this monotonicity. But we can argue as follows: Assume $L_0(A \cup B) = 0$. Then $U_0(\neg A \cap \neg B) = 1$, thus by (21), $U(\neg A \cap \neg B) = 1$. Using (17), we get $U(\neg A) = U_0(\neg A \cap \neg B) \cdot U(\neg A) \leq U(\neg A \cap \neg B) \cdot U_0(\neg A) = U_0(\neg A)$, contradicting (22).

[13] Concerning the modified proof of model 2 mentioned above, note that due to (16) this case does not occur, if $\varepsilon = 0$.

6 Concluding Remarks

To start with a topic raised in Section 1, consider again equation (2), that is $L = f \circ L_0$, with the standard assumption that $f\colon [0;1] \to [0;1]$ is a function with $f(0) = 0$ and $f(1) = 1$. Already in [2], Proposition 5.2, it is shown that via (2) every *convex* f transfers any given F-field $\mathcal{O}_0 = (\Omega; \mathcal{A}; L_0)$ to a neighbourhood $\mathcal{O} = (\Omega; \mathcal{A}; L)$, which is an F-field too. But in a strict sense, this neighbourhood model is not "appropriate", since by Theorem 1, model 3, there is a weaker condition on f doing the same — namely the condition of bi-elasticity. The question arises, whether this requirement is "appropriate" instead. Indeed, bi-elasticity is even the weakest assumption on f ensuring that via (1), i.e. $L = f \circ p$, every K-function $p \in \mathcal{K}(\Omega; \mathcal{A})$ is transfered to an F-neighbourhood $\mathcal{O} = (\Omega; \mathcal{A}; L)$, if we are allowed to vary the underlying measurable space $(\Omega; \mathcal{A})$. For this, there is a quick argument, if additionally it is assumed that our functions f are continuous on $]0;1[$. In this case it is even sufficient to consider all *finite* measurable spaces and, for each of them, only *one* central distribution p "testing" equation (1):

Let f be fixed, being continuous on $]0;1[$ and having the above-mentioned property of generating F-neighbourhoods via (1). We restrict ourselves in deriving condition (4), where, by continuity, it is possible to assume that in there y and z are rational numbers: $y = \frac{i}{k}$ and $z = \frac{j}{k}$ for $0 < i \le j \le k$. Now, for this $k \in \mathbb{N}$, we walk up to the finite measurable space $(\Omega_k; \mathcal{P}(\Omega_k))$ and employ the corresponding classical uniform probability $p = p_0^k$ as central distribution, generating via (1) an F-neighbourhood $\mathcal{O} = (\Omega_k; \mathcal{P}(\Omega_k); L)$ (cf. Example 1). Hence $\mathcal{O}$ is an uniform F-field on $(\Omega_k; \mathcal{P}(\Omega_k))$, which — according to the last sentence in Example 1 — obeys the chain (14), especially its left part. Finally, an easy transformation leads to the desired inequality in (4).

Apart from this — last — positive result given here, many questions concerning the role of bi-elasticity within the theory of interval probability remain open. For example, the *concept of conditional interval probability* is still debated (see [10] or [12] and the references therein). In particular, Weichselberger's notion of the *canonical concept* has the disadvantage that some constructions are not closed w.r.t. the F-property. Using the corresponding notation $\Psi(A \mid B) = \frac{\Psi(A)}{\Psi(B)}$ for every $A, B \in \mathcal{A}$ with $A \subseteq B$ such that $\Psi(B) \neq 0$, where Ψ can be a K-function as well as the lower or the upper bound of a probability field, it is possible to rephrase sensitively Theorem 3, models 1 and 2, and Theorem 4. Considering the outcome, it perhaps is feasible to modify these theorems in a way, which is profitable for a better understanding of the phenomenon of conditional interval probability.

Summarizing, the results presented in this article can be seen as the formal basis for joining together robust statistics and interval probability in its most expressive form, i.e., the concept of coherent or F-probabilities. The systematic development of distorted probabilities should be able to initiate a variety of applications in robust statistics and beyond.

Acknowledgements

I thank Thomas Augustin and three anonymous referees for their valuable remarks.

References

[1] T. Augustin. *Optimale Tests bei Intervallwahrscheinlichkeit*. Vandenhoeck & Ruprecht, Göttingen, 1998.

[2] T. Augustin. 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*, 105: 149–173, 2002.

[3] A. Buja. On the Huber-Strassen theorem. *Probability Theory and Related Fields*, 73: 149–152, 1986.

[4] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer, Dordrecht, 1994.

[5] D. Greenwald. *The McGraw-Hill Dictionary of Modern Economics*. McGraw-Hill Book Company, New York, 1973.

[6] P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[7] P.J. Huber and V. Strassen. Minimax tests and the Neyman-Pearson lemma for capacities. *Annals of Statistics*, 1: 251–263, 1973. Correction: 2: 223–224, 1974.

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

[9] A. Wallner. *Beiträge zur Theorie der Intervallwahrscheinlichkeit — Der Blick über Kolmogorov und Choquet hinaus*. Dr. Kovač, Hamburg, 2002.

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

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

[12] K. Weichselberger and T. Augustin. *On the Symbiosis of Two Concepts of Conditional Interval Probability*. `www.stat.uni-muenchen.de/~thomas/weichselberger-augustin-isipta03long.pdf`, 2003.

Anton Wallner is with the Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstr. 33, 80539 München, Germany. E-mail: toni@stat.uni-muenchen.de