BELIEF FUNCTIONS

ON INFINITE DOMAINS

HUNG T. NGUYEN

New Mexico State University (USA)

and

Chiang Mai University (THAILAND)

CARTHAGE, MAY 2013

OBJECTIVE AND OUTLINE OF THE LECTURE

Presenting an **theoretical** introduction to the extension of the theory of belief functions to infinite domains.

Details of the presented material are given in another handout.

1. Belief Functions: From finite domains to infinite domains

2. Random closed sets

3. Capacity functionals

4. Multivariate capacity functionals

1. BELIEF FUNCTIONS

Let U be a set. According to Shafer, a belief function (as a mathematical representation of an evidence) is a function $F : 2^U \longrightarrow [0, 1]$ such that

(i) $F(\varnothing) = 0$, $F(U) = 1$

(ii) Monotonicity of infinite order: for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in 2^U,

$$F(\cup_{j=1}^{k} A_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} F(\cap_{i \in I})$$

Examples

a) Let U be a *finite* set. Let X be a random element, defined on some $(\Omega, \mathcal{A}, P)$ with values in 2^U such that $P(X = \varnothing) = 0$. Then $F(A) = P(X \subseteq A)$ is a belief function on the *finite* domain 2^U .

b) Let $(U, \mathcal{U}, P)$ be an arbitrary probability space, e.g., $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), dH)$, where H is some multivariate distribution function on $\mathbb{R}^d$. Then $F : 2^U \longrightarrow [0, 1]$, defined by

$$F(A) = \sup\{P(B) : B \in \mathcal{U}, B \subseteq A\}$$

is a belief function on the *infinite* domain 2^U.

When U is **finite**, the above example a) contains all belief functions: F is a belief function on a finite set U if and only if it is the distribution of some *nonempty random set* X.

This observation is useful since, instead of viewing belief functions as generalizations of probability measures (as set functions), we can view them as distribution functions, not of single-valued random elements (e.g., random vectors), but of set-valued ones (namely, random sets), so that the umbrella of our studies remains entirely within the well-established theory of probability. This has became clear as far as the important problem of combination of evidence is concerned.

Because of the above fact, an "equivalent" definition to the original *axiomatic* definition of Shafer is this. Let $f : 2^U \rightarrow [0, 1]$ such that $f(\varnothing) = 0$, and $\sum_{A \subseteq U} f(A) = 1$ (which is a bona fide probability density, on 2^U, of some finite random set). Then $F(A) = \sum_{B \subseteq A} f(B)$ is the associated belief function. The fact that belief functions on finite

domains have "densities" does not extend to infinite domains. Thus, axiomatic definitions are still needed in general.

Remark: the existence of densities in the finite domain case is due to the fact that $(2^U, \subseteq)$ is a (locally) finite poset. The Mobius inverse of the set function F is a probability density on 2^U in view of the axioms of belief functions (a remarkable fact). The situation is different for non locally finite posets, althought it is possible to extend Carlos Rota's Mobius transforms to them, as in M. Marinacci (1996) " Decomposition and representation of coalitional games", *Math. Oper. Res. 21, 1000-1015*.

In other words, the original axiomatic definition of belief functions of Shafer which is valid for any domain, finite or not, should be kept in mind, especially when considering infinite domains.

The axiomatic definition of belief functions (distributions of random sets) is the analogue of that of distribution functions of real-valued random variables. Distribution functions might have densities or not. They are defined axiomatically. For example, multivariate distributions are axiomatically defined differently than univariate ones. They correspond uniquely to probability measures (Lebesgue-Stieltjes theorem).

As a prelude to multivariate belief functions on infinite domains, let's say few words about **multivariate belief functions on finite domains**, for the purpose of combination of evidence, for example.

Let U, V be two finite sets. What is a belief function on $2^U \times 2^V$?

Not extending the axiomatic definition in the univariate case, but imitating the univariate case, you will say, well, it is a function $H : 2^U \times 2^V \longrightarrow [0, 1]$ built up by a bivariate probability density $h : 2^U \times 2^V \longrightarrow [0, 1]$ as $H(A, B) = \sum_{A' \subseteq A, B' \subseteq B} h(A', B')$.

We need to be rigorous and consistent here!

As we will see, in the problem of combination of evidence, a joint belief function plays the role of a joint distribution of a random set vector which is well defined in probability theory.

Here is the **axiomatic definition of multivariate belief functions**, due to B. Schmelzer (2012), "Characterizing joint distributions of random sets my multivariate capacities", *IJAR (53), 1228-1247*, that I give for the case of two belief functions, the formulation for several belief functions is a straightforward extension. It is the analogue of the axiomatic definition of multivariate distributions of random vectors as compared to that of univariate distributions. It is the extension of Shafer's definition of univariate belief functions to higher dimensions, in the sense that the axioms of, say, of a bivariate belief function H on $2^U \times 2^V$ makes it a joint distribution of a nonempty random set vector.

Definition of a bivariate (joint) belief function

A joint belief function (on a finite domain) is a function $H : 2^U \times 2^V \rightarrow [0, 1]$ such that

(i) $H(\varnothing, \varnothing) = 0$, $H(U, V) = 1$

(ii) $H(.,.)$ is jointly monotone of infinite order: for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in 2^U; $B_1, B_2, ..., B_k$ in 2^V,

$$H(\cup_{j=1}^{k} A_j, \cup_{j=1}^{k} B_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} H(\cap_{i \in I} A_i, \cap_{i \in I} B_i)$$

The upshot is this.

Theorem (Schmelzer, 2012)

H is a joint belief function if and only if it is of the form $H(A, B) = P(X \subseteq A, Y \subseteq B)$, where X, Y are random sets, defined on $(\Omega, \mathcal{A}, P)$ with $P(X = \varnothing, Y = \varnothing) = 0$.

Just like in the univariate case, the axioms of a joint belief function H makes the Mobius inverse h of H a bona fide joint probability density. Specifically,

$$h(A, B) = \sum_{(A', B') \leq (A, B)} (-1)^{|A \setminus A, \setminus'| + |B \setminus B'|} H(A', B')$$

Perhaps, it is here that we need to bring in **incidence algebras** as a useful computational tool in the study of belief functions.

INCIDENCE ALGEBRAS

In combinatorics, a partially ordered set (poset) $(S, \preceq)$ is said to be locally finite if every "interval" $\{x : a \preceq x \preceq b\}$ is finite.

The set $\mathcal{V} = \{f : 2^U \to \mathbb{R}\}$, with pointwise addition and multiplication $r \in \mathbb{R}$, $rf(.)$, is a vector space over $\mathbb{R}$, with dimension $2^{|U|}$.

The set $\mathcal{W} = \{\alpha : (2^U)^{[2]} \to \mathbb{R}\}$, where

$$(2^U)^{[2]} = \{(A, B) : A \subseteq B \subseteq U\}$$

with pointwise addition and multiplication defined by

$$(\alpha * \beta)(A, B) = \sum_{A \subseteq Z \subseteq B} \alpha(A, Z)\beta(Z, B)$$

is called the *incidence algebra* of U over the field $\mathbb{R}$. It is a ring with identity which is the function on $\mathcal{W}$ defined by

$$\delta(A, B) = \begin{cases} 1 & \text{for} \quad A = B \\ 0 & \text{for} \quad A \neq B \end{cases}$$

For an element $\alpha \in \mathcal{W}$ to have an inverse, it is necessary and sufficient that for all A, $\alpha(A, A) \neq 0$, and then its inverse is given inductively by

$$\alpha^{-1}(A, A) = \frac{1}{\alpha(A, A)}$$

$$\alpha^{-1}(A,B) = [\frac{-1}{\alpha(A,A)}] \sum_{A \subseteq Z \subseteq B} \alpha(A,Z)\alpha^{-1}(Z,B)$$

for $A \subseteq B$.

Elements that have an inverse are called *units*. There are two very special units in $\mathcal{W}$:

(1) The *Mobius function* $\mu(A,B) = (-1)^{|B \setminus A|}$ (for $A \subseteq B$, and 0 otherwise),

(2) The *Zeta function* $\xi(A,B) = 1$ (for $A \subseteq B$, and 0 otherwise).

which are inverses of each other, i.e., $\mu * \xi = \xi * \mu = \delta$

There is a natural operation on the elements of the vector space $\mathcal{V}$ by the elements of the incidence algebra $\mathcal{W}$: For $f \in \mathcal{V}$, and $\alpha \in \mathcal{W}$, define $f * \alpha : 2^U \to \mathbb{R}$ by

$$(f * \alpha)(A) = \sum_{Z \subseteq A} f(A)\alpha(Z, A)$$

In particular,

$$(f * \mu)(A) = \sum_{Z \subseteq A} f(A)\mu(Z, A) = \sum_{Z \subseteq A} (-1)^{|A \setminus Z|} f(Z)$$

is called the *Mobius inverse* of f or the Mobius inversion of f.

Thus, the probability density function (or "mass" function)

$$f(A) = (F * \mu)(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

is the Mobius inverse of the belief function F.

The maps $f \in \mathcal{V} \rightarrow f * \mu \in \mathcal{V}$ and $f \rightarrow f * \xi$ are one to one maps from $\mathcal{V}$ onto $\mathcal{V}$, and are inverses of one another. Specifically, let $\varphi(f) = f * \mu$, and $\psi(f) = f * \xi$, then $\varphi \circ \psi = \psi \circ \varphi = id$. Indeed,

$$(\varphi \circ \psi)(f)(A) = \varphi[\psi(f)](A) = \varphi[f * \xi](A) = [(f * \xi) * \mu](A)$$

$$= [f * (\xi * \mu)](A) = (f * \delta)(A) = \sum_{Z \subseteq A} f(Z)\delta(Z, A) = f(A)$$

Clearly, $f = F * \mu \Longleftrightarrow F = f * \xi$. Indeed, if $f = F * \mu$, then $f * \xi = F * \mu * \xi = F * \delta = F$; if $F = f * \xi$, then $F * \mu = f * \xi * \mu = f * \delta = f$.

For

$$F(A) = P(S \subseteq A) = \sum_{B \subseteq A} f(B) = (f * \xi)(A)$$

we have

$$f(A) = (F * \mu)(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

Similarly, for

$$f(A) = (F * \mu)(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

we get

$$F(A) = (f * \xi)(A) = \sum_{B \subseteq A} f(A)\xi(B, A) = \sum_{B \subseteq A} f(B)$$

This is referred to as *Mobius inversion formula*.

Uncertainty measures, such as probability measures, are defined axiomatically. For belief functions, viewed as uncertainty measures, the axioms (i) and (ii) are natural as far as the intuitive concept of belief

is concerned. The axiom (iii), the monotonicity of infinite order, is interesting since it corresponds precisely (and uniquely) to the concept of distribution of random sets. Specifically, let μ be the Mobius function on the (finite) poset $(2^U, \subseteq)$, and $F : 2^U \to [0, 1]$. Then F is a belief function if and only if $F * \mu$ is a probability density f on 2^U with $f(\varnothing) = 0$.

As we will see, in multivariate belief functions, it is convenient to consider Mobius function on product of locally finite partial ordered sets (locally finite posets).

Let U and V be two finite sets, then the product $2^U \times 2^V$ is a (finite) poset with partial order defined by $(A, B) \leq (A', B')$ iff $A \subseteq A'$, and $B \subseteq B'$. If μ denotes the Mobius function (which is integer-valued and unique) of $(2^U, \subseteq)$, then the Mobius function ν of $(2^U \times 2^V, \leq)$ is

$$\nu((A, B), (A', B')) = \mu(A, A')\mu(B, B') = (-1)^{|A' \setminus A| + |B' \setminus B|}$$

An example

Let F, G be two belief functions on the same finite domain U. We seek a possible joint belief function H with marginals F, G, i.e., $H : 2^U \times 2^U \to [0,1]$ such that $F(A) = H(A, U)$, $G(B) = H(U, B)$.

In terms of densities, f, g, h, respectively, we can proceed as follows.

Consider the support $\mathcal{F}(f) = \{A \subseteq U : f(A) > 0\}$ of finite random sets taking values in 2^U with density f.

Let $A_1, A_2, ..., A_m$ a listing of the elements of $\mathcal{F}(f)$. Take $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}_1, \lambda(dx))$ and partition $[0,1]$ into m intervals I_i with $\lambda(I_i) = f(A_i)$. Define the random set $X : [0,1] \to 2^U$ by $X(x) = A_i$ iff $x \in I_i$, then X is a (canonical) random set representation of F, i.e., $F(A) = P(X \subseteq A)$.

Similarly for G: $B_1, B_2, ..., B_n$ is a given listing of elements of $\mathcal{F}(g)$ and $[0,1] = \cup_{j=1}^{n} J_j$, with $\lambda(J_j) = g(B_j)$, $Y : [0,1] \to 2^U : Y(y) = B_j$ iff $y \in J_j$, $G(B) = P(Y \subseteq B)$.

Since the random set vector (X, Y) is defined on the product space $([0,1]^2, \mathcal{B}_1 \otimes \mathcal{B}_1)$ with marginal probability measures corresponding to uniform distributions on $[0,1]$, any "joint" probability measure is of the form dC where C is a copula (C is a bivariate distribution function on $[0,1]^2$ with uniform marginals, dC is its associated probability measure on $\mathcal{B}_1 \otimes \mathcal{B}_1$). Thus, one possible joint belief function is

$$H^C_{(X,Y)}(A, B) = dC\{(x, y) \in [0,1]^2 : X(x) \subseteq A, Y(y) \subseteq B\}$$

whose density is

$$h^C_{(X,Y)}(A,B) = \sum_{(A',B')\leq(A,B)} (-1)^{|A\setminus A'|+|B\setminus B'|} H^C_{(X,Y)}(A',B')$$

From which, a belief function for $X \cap Y$ on 2^U can be taken as conditional distribution of $X \cap Y | X \cap Y \neq \varnothing$, where the density of $X \cap Y$ is given by

$$dC(X \cap Y = A) = \sum_{B\cap D=A} h^C_{(X,Y)}(B,D)$$

CAPACITY FUNCTIONALS

A dual notion to belief function $F(.)$ is the set function $T : 2^U \rightarrow [0, 1]$, defined by

$$T(A) = 1 - F(A^c)$$

called *a capacity functional.*

From the axioms of belief functions, it follows that $T(.)$ has the properties:

$$T(\varnothing) = 1 - F(U) = 0, \quad T(U) = 1 - F(\varnothing) = 1$$

and for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in 2^U,

$$T(\cap_{j=1}^{k} A_j) = 1 - F(\cup_{j=1}^{k} A_j^c) \leq 1 - \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i^c)$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} T(\cup_{i \in I} A_i)$$

Thus, axiomatically, a capaxcity functional $T(.)$ is defined as:

A function $T : 2^U \rightarrow [0, 1]$ is called a capacity functional on a finite set U, if it satisfies

(i) $T(\varnothing) = 0$

(ii) $T(U) = 1$

(ii) T is *alternating of infinite order, i.e.,*

For any $k \geq 2$, and $A_1, A_2, ..., A_k$ in 2^U,

$$T(\cap_{j=1}^{k} A_j) \leq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} T(\cup_{i \in I} A_i)$$

Remark

Dually, given a capacity functional T, the associated belief function is $F(A) = 1 - T(A^c)$. Since there exist a probability space $(\Omega, \mathcal{A}, P)$ and a nonempty random set $X : \Omega \to 2^U$ such that $F(A) = P(X \subseteq A)$, we have

$$T(A) = 1 - F(A^c) = 1 - P(X \subseteq A^c) = P(X \cap A \neq \varnothing)$$

If we let $\mathcal{F}_A = \{B \subseteq U : B \cap A \neq \varnothing\}$, and Q be the probability law of the random set X, i.e., $Q = PX^{-1}$, which is the probability measure of 2^{2^U} determined by the density $f : 2^U \to [0, 1]$ of F (its Mobius inverse): for $\mathbb{A} \in 2^{2^U}$, $Q(\mathbb{A}) = \sum_{A \in \mathbb{A}} f(A)$, then $T(A) = Q(\mathcal{F}_A)$.

It is this relation which is convenient to use in the case of infinite domains.

Example

Let $\pi : U \to [0,1]$ with $\max_{u \in U} \pi(u) = 1$, and $\alpha : \Omega \to [0,1]$ be a random variable, uniformly distributed.

Let: $X : \Omega \to 2^U$,

$$X(\omega) = \{u \in U : \pi(u) \geq \alpha(\omega)\}$$

The capacity functional of X is, for $A \neq \varnothing$, is

$$T(A) = P(X \cap A \neq \varnothing) = P(\omega : \max_{u \in A} \pi(u) \geq \alpha(\omega)) = \max_{u \in A} \pi(u)$$

The associated belief function is

$$F(A) = 1 - T(A^c) = 1 - \max_{u \in A^c} \pi(u)$$

and the probability density of X is

$$f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} [1 - \max_{u \in A^c} \pi(u)]$$

Note that the function $\pi(.)$ is precisely the *covering function* of the random set X, i.e.,

$$\pi(u) = P(X \cap \{u\} \neq \varnothing\} = P(u \in X)$$

2. RANDOM CLOSED SETS

We consider now the case where the domain U is infinite. To have a "beautiful" theory extending the random set representation of belief functions in the finite case, we restrict ourself to the concrete and practical cases of euclidean spaces $\mathbb{R}^d$, where a theory of *random closed sets* has fully developped (see G. Matheron, 1975).

Remark: A more general framework for studying belief functions on *fuzzy sets*, i.e., *random fuzzy sets*, can be carried out within the theory of *Continuous Lattices* (G. Gierz et al, 2003, *Continuous Lattices and Domains*, Cambridge Univ. Press), where the space of u.s.c. functions (fuzzy closed sets), equiped with $\supseteq$, is a continuous lattice, in view of the fact that the hit-or-miss topology on the the space $\mathcal{F}(\mathbb{R}^d)$ of closed sets of $\mathbb{R}^d$ is the Lawson topology. See H.T. Nguyen and H. Tran (2007), On a continuous lattice approach to modeling of coarse data in systems anslysis, *Journal of Uncertain Systems, 1(1), 62-73.*

To be concrete, we consider $U = \mathbb{R}^d$, although the following exposition is valid for more general Hausdorff, locally compact and second countable spaces.

Random sets in the case of U finite are in fact random closed sets (i.e., random element taking closed sets as values) in it discrete topology. Thus, it is natural to use the existing theory of random closed sets to study belief functions on $\mathbb{R}^d$.

The big picture is this. Belief functions are viewed as **distributions of random closed sets** via containment functionals, the natural analog of cumulative distributions. Such containment functionals are defined to be belief functions. Note that, for any set U (finite or not), if we consider its discrete topology (every subset of it is open), then U is a topological space which is Hausdorff and locally compact, and the containment $A \in 2^U \rightarrow P(X \subseteq A)$ determines uniquely the probability law of X on 2^{2^U}: $P(X \in \mathbb{A})$ for any $\mathbb{A} \in 2^{2^U}$.

As stated, we will be using the dual $T(A) = 1 - F(A^c)$, in view of the celebrated Choquet theorem, characterizing distributions of random closed sets by capacity functionals, the analog of Lebesgue-Stieltjes theorem of real-valued random variables.

As we will see, the axiomatic definition of belief functions in the infinite domain case is a natural extension of that in the finite domain case, where continuity (in some topology) will be called upon.

First, let's recall the basics of random closed sets on $\mathbb{R}^d$.

To define a random element X, defined on some probability space $(\Omega, \mathcal{A}, P)$ with values in $\mathcal{F}(\mathbb{R}^d)$, the set of closed sets of $\mathbb{R}^d$, we need to specify a $\sigma-$ field of subsets of $\mathcal{F}(\mathbb{R}^d)$.

The natural way for this is to define a topology of $\mathcal{F}(\mathbb{R}^d)$ and take its Borel $\sigma-$ field.

Let $\mathcal{F}, \mathcal{G}$ and $\mathcal{K}$ be the classes of closed, open and compact subsets of $\mathbb{R}^d$, respectively.

Following Matheron (1975), we will take $\mathcal{B}(\mathcal{F})$ as the Borel $\sigma-$ field generated by the hit-or-miss topology $\mathcal{T}$ on $\mathcal{F}$ which is constructed as follows. For $A \subseteq \mathbb{R}^d$, let

$$\mathcal{F}_A = \{F \in \mathcal{F} : F \cap A \neq \varnothing\}$$

$$\mathcal{F}^A = \{F \in \mathcal{F} : F \cap A = \varnothing\}$$

$$\mathcal{F}^K_{G_1,G_2,...,G_n} = \mathcal{F}^K \cap \mathcal{F}_{G_1} \cap ... \cap \mathcal{F}_{G_n}$$

Let

$$\mathbb{B} = \{\mathcal{F}^K_{G_1,G_2,...,G_n} : K \in \mathcal{K}, G_i \in \mathcal{G}, i = 1, 2, ..., n\}$$

Note that in $\mathbb{B}$, when $n = 0$, $\mathcal{F}^K_{G_1,G_2,...,G_n}$ stands for $\mathcal{F}^K$. Also, $\mathcal{F}_G = \mathcal{F}^{\varnothing}_G$, $\mathcal{F}_{\varnothing} = \varnothing$, $\mathcal{F} = \mathcal{F}^{\varnothing}$.

The collection $\mathbb{B}$ is a basis for a topology $\mathcal{T}$ which is called the hit-or-miss topology on $\mathcal{F}$. Its Borel $\sigma-$ field is denoted as $\mathcal{B}(\mathcal{F})$.

By a random closed set we mean a $\mathcal{A} - \mathcal{B}(\mathcal{F})$ map $X : \Omega \to \mathcal{F}$. Its probability law is the probability measure $P_X = PX^{-1}$ on $\mathcal{B}(\mathcal{F})$.

Examples

(1) Let $\varphi : \mathbb{R}^d \to [0, 1]$ be upper semi-continuous (u.s.c.), i.e., $\{x \in \mathbb{R}^d : \varphi(x) \geq a\} \in \mathcal{F}$ for any $a \in \mathbb{R}$. Let $\alpha : \Omega \to [0, 1]$ be a random variable, uniformly distributed on $[0, 1]$. Define $S : \Omega \to \mathcal{F}$ by

$$S(\omega) = \{x \in \mathbb{R}^d : \varphi(x) \geq \alpha(\omega)\}$$

(2) Let $S : \Omega \to \mathcal{F}$ be a random closed on $\mathbb{R}^d$, and $\lambda(dx)$ be the Lebesgue measure on $\mathcal{B}(\mathbb{R}^d)$. Then $\lambda(S) : \Omega \to \mathbb{R}^+$ is a random variable. Indeed, let $\psi : \mathbb{R}^d \times \mathcal{F} \to \{0, 1\}$ be

$$\psi(x, F) = \begin{cases} 1 & \text{if} \quad x \in F \\ 0 & \text{if} \quad x \notin F \end{cases}$$

Then the restriction of $\lambda(dx)$ to $\mathcal{F}$ is

$$\lambda(F) = \int_{\mathbb{R}^d} \psi(x, F)\lambda(dx)$$

so that, by Fubini Theorem, $\lambda : \mathcal{F} \to \bar{\mathbb{R}}$ is $\mathcal{B}(\mathcal{F})-$ $\mathcal{B}(\bar{\mathbb{R}})-$ measurable, and hence $\lambda(S) = \lambda \circ S$ is $\mathcal{A} - \mathcal{B}(\bar{\mathbb{R}})-$ measurable.

3. CAPACITY FUNCTIONALS

Recall that probability measures on $\mathcal{B}(\mathbb{R}^d)$ is characterized by distribution functions, defined on $\mathbb{R}^d$ (Lebesgue-Stieltjes theorem). The beauty of random closed sets is that a similar characterization exists (Choquet theorem): Probability measures on $\mathcal{B}(\mathcal{F})$ are characterized by set functions, defined on $\mathcal{K}$.

Recall that, in the case of a finite set U, the domain of a capacity functional T is the whole power set 2^U: Let $T : 2^U \to [0, 1]$ be such that

(i) $T(\varnothing) = 0$

(ii) T is alternating of infinite order, i.e., for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in 2^U,

$$T(\cap_{j=1}^{k} A_j) \leq \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} T(\cup_{i \in I} A_i)$$

Then T characterizes a probability measure Q on 2^{2^U}, i.e., a probability law of some random set S on U, via

$$T(A) = Q\{B \subseteq U, B \cap A \neq \varnothing) = Q(\mathcal{F}_A)$$

More specifically, there exist a probability space $(\Omega, \mathcal{A}, P)$, a random element $S : \Omega \to 2^U$ such that, for any $A \in 2^U$,

$$P(S \cap A \neq \varnothing) = P_S(\mathcal{F}_A) = Q(\mathcal{F}_A) = T(A)$$

In the case of $\mathbb{R}^d$, what should be the domain for corresponding capacity functionals? The following motivates the choice of $\mathcal{K}$.

Random vectors, i.e., random elements taking values in $\mathbb{R}^d$, can be viewed as special random closed sets on $\mathbb{R}^d$ which take singleton sets as values. Specifically, if $X : \Omega \to \mathbb{R}^d$ is a random vector, then $S = \{X\}$ is a random closed set on $\mathbb{R}^d$, and we have, for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P_X(A) = P(X \in A) = P(S \cap A \neq \varnothing)$$

However, the probability law P_X of X, defined on $\mathcal{B}(\mathbb{R}^d)$, is determined from the values of P_X on the class $\mathcal{K}$ of compact subsets of $\mathbb{R}^d$. Specifically, for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P_X(A) = \sup\{P_X(K) : K \in \mathcal{K}, \quad K \subseteq A\}$$

From the above fact, we see that it suffices to consider P_X on $\mathcal{K} \subseteq \mathcal{B}(\mathbb{R}^d)$.

If we let $T : \mathcal{K} \to [0, 1]$ where

$$T(K) = P(X \in K) = P_X(K)$$

then the set function $T(.)$ determines completely the probability law P_X of the random vector X. It can be verified that, $T(.)$ has the following properties

(i) $T(\varnothing) = 0$

(ii) If $K_n \in \mathcal{K}$ such that $K_n \searrow K$ (i.e., $K_n \subseteq K_{n+1}$, $K = \cap_{n\geq 1} K_n$), then $T(K_n) \searrow T(K)$

(iii) T is alternating of infinite order, i.e., for $n \geq 2$, and $K_1, K_2, ..., K_n$ in $\mathcal{K}$,

$$T(\cap_{j=1}^{n} K_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} T(\cup_{i\in I} K_i)$$

Now let $S = \{X\}$ be a singleton-valued random closed set on $\mathbb{R}^d$, we have

$$P_X(K) = P(X \in K) = P(\{X\} \cap K \neq \varnothing) = P(S \cap K \neq \varnothing)$$

Thus, for a general random closed set S on $\mathbb{R}^d$, we consider the set function $T_S : \mathcal{K} \to [0, 1]$ defined by

$$T_S(K) = P(S \cap K \neq \varnothing) = P_S(\mathcal{F}_K)$$

Definitions

(1) The set function $T_S : \mathcal{K} \to [0, 1]$ defined by

$$T_S(K) = P(\omega : S(\omega) \cap K \neq \varnothing) = P_S(\mathcal{F}_K)$$

is called the **capacity functional** of S.

(2) A set function $T : \mathcal{K} \to [0, 1]$ such that

(i) $T(\varnothing) = 0$

(ii) If $K_n \in \mathcal{K}$ such that $K_n \searrow K$ (i.e., $K_n \subseteq K_{n+1}$, $K = \cap_{n \geq 1} K_n$), then $T(K_n) \searrow T(K)$

(iii) T is alternating of infinite order, i.e., for $n \geq 2$, and $K_1, K_2, ..., K_n$ in $\mathcal{K}$,

$$T(\cap_{j=1}^{n} K_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} T(\cup_{i \in I} K_i)$$

is called a *capacity functional.*

Just like the special case of $S = \{X\}$, the capacity functional T_S of any random closed set S on $\mathbb{R}^d$ satisfies the above three basic properties (i),(ii) and (iii).

The upshot is that these properties characterize probability laws of random closed sets.

Choquet Theorem

If $T : \mathcal{K} \to [0, 1]$ is a capacity functional, then T is the capacity functional of some random closed set S on $\mathbb{R}^d$. Specifically, there exists a unique probability measure Q on $\mathcal{B}(\mathcal{F})$ such that, for all $K \in \mathcal{K}$, $Q(\mathcal{F}_K) = T(K)$.

Remark

(1) The capacity functional T, defined on $\mathcal{K}$, can be extended to the power set of $\mathbb{R}^d$, preserving the above characteristic properties, as follows.

For $G \in \mathcal{G}$, set

$$T(G) = \sup\{T(K) : K \subseteq G, K \in \mathcal{K}\}$$

then, for $A \subseteq \mathbb{R}^d$,

$$T(A) = \inf\{T(G) : G \supseteq A, G \in \mathcal{G}\}$$

(2) The dual of T is referred to as a belief function on $\mathbb{R}^d$: $A \rightarrow 1 - T(A^c)$.

Example

Let $\varphi : \mathbb{R}^d \to [0,1]$ be u.s.c. and $T(K) = \sup_{x \in K} \varphi(x)$. Then, $T(.)$ is a capacity functional.

While a probabilistic proof is simple, we verify the basic properties of $T(.)$ in this example directly.

Since the supremum over the empty set is zero, $T(\varnothing) = 0$. $T(.)$ is alternating of infinite order since it is a maxitive set function. Finally, the continuity from above of $T(.)$ is seen as follows.

Let $K_n \searrow K$ in $\mathcal{K}$. We have

$$T(K) \leq \inf_n T(K_n) = s$$

Let $\varepsilon > 0$ and consider

$$A_n = \{x \in \mathbb{R}^d : \varphi(x) \geq s - \varepsilon\} \cap K_n$$

By hypothesis, the set $\{x \in \mathbb{R}^d : \varphi(x) \geq s - \varepsilon\}$ is closed, and so is A_n. The A_n are contained in K_1, and hence $\cap_{n \geq 1} A_n \neq \varnothing$ (since the A_n have the finite intersection property, and noting that $A_n \neq \varnothing$ for each n). Since $A_n \subseteq K_n$ for all n, it follows that

$$T(K) \geq \sup_{x \in \cap_{n \geq 1} A_n} \varphi(x) \geq s - \varepsilon$$

implying that $\lim_{n \to \infty} T(K_n) = T(K)$.

4. MULTIVARIATE CAPACITY FUNCTIONALS

Just like the case of finite domains, we consider the setting of joint distributions of random closed set *vectors* in order to investigate the problem of combination of evidence on infinite domains. The recent work of Schmelzer (2012) seems to be the best presentation of the issue.

Again, for simplicity, we address the case of two random closed sets X and Y on the same domain $\mathbb{R}^d$. They are defined on a probability space $(\Omega, \mathcal{A}, P)$ with values in $\mathcal{F}(\mathbb{R}^d)$. We consider the vector $(X, Y) : \Omega \rightarrow \mathcal{F} \times \mathcal{F} : (X, Y)(\omega) = X(\omega) \times Y(\omega)$.

The $\sigma-$field on $\mathcal{F} \times \mathcal{F}$ is $\mathcal{B}(\mathcal{F}) \otimes \mathcal{B}(\mathcal{F})$ which is generated by sets $\mathcal{F}_K \times \mathcal{F}_{K'}, K, K' \in \mathcal{K}$.

If we let $P_{(X,Y)}$ be the probability law of (X, Y) on $\mathcal{B}(\mathcal{F}) \otimes \mathcal{B}(\mathcal{F})$, then

$$\begin{aligned} P_{(X,Y)}(\mathcal{F}_K \times \mathcal{F}_{K'}) &= P((X \times Y) \cap (K \times K') \neq \varnothing) \\ &= P(X \cap K \neq \varnothing, Y \cap K' \neq \varnothing) = P(X^-(K) \cap Y^-(K') \end{aligned}$$

where we set $X^-(K) = \{\omega : X(\omega) \cap K \neq \varnothing\}$.

We identify each (K, K') with their co-product $K \amalg K' = (K \times \{1\}) \cup (K' \times \{2\})$ which is a subset of $\mathcal{K} \amalg \mathcal{K}$. Similarly, we identify $\mathcal{F} \times \mathcal{F}$ with their co-product $\mathcal{F} \amalg \mathcal{F} = (\mathcal{F} \times \{1\}) \cup (\mathcal{F} \times \{2\})$. In this way, the vector (X, Y) is identified with the random set $S(\omega) = X(\omega) \amalg Y(\omega)$ on $\mathcal{F} \amalg \mathcal{F}$. Note that

$$\{(X \amalg Y) \cap (K \amalg K') \neq \varnothing\} = \{X^{-}(K) \cup Y^{-}(K')\}$$

Thus, the (**multivariate**) capacity functional of (X, Y) can be written as

$$T_{(X,Y)} : \mathcal{K} \times \mathcal{K} \to [0, 1] : T_{(X,Y)}(K, K') = P(X^{-}(K) \cup Y^{-}(K'))$$

It can be verified that

(i) $T_{(X,Y)}(\varnothing, \varnothing) = 0$

(ii) jointly alternating of infinite order, i.e., for any $k \geq 2$, and $K_j, K'_j, j = 1, 2, ..., k$ compacts,

$$T(\cap_{j=1}^{k} K_j, \cap_{j=1}^{k} K'_j) \leq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} T(\cup_{i \in I} K_i, \cup_{i \in I} K'_i)$$

(iii) jointly continuous from above, i.e., for any decreasing sequences of compacts K_n, K'_n, we have

$$\lim_{n \to \infty} T(K_n, K'_n) = T(\cap_{n \geq 1} K_n, \cap_{n \geq 1} K'_n)$$

As expected, a multivariate version of Choquet Theorem is this.

Given a multivariate capacity functional T, defined on $\mathcal{K} \times \mathcal{K}$, we define the associated set function on $\mathcal{K} \amalg \mathcal{K}$ by $T^*(K \amalg K') = T(K, K')$. Then, T^* is a (univariate) capacity functional if and only if T is a multivariate capacity funtional.

Multivariate Choquet Theorem (Schmelzer, 2012)

Let T be a multivariate capacity functional on $\mathcal{K} \times \mathcal{K}$. Then there exists a unique probability measure Q on $\mathcal{B}(\mathcal{F} \times \mathcal{F})$ such that $T(K, K') = Q(\mathcal{F}_K^* \cup \mathcal{F}_{K'}^*)$, for all K, K' in $\mathcal{K}$ and where $\mathcal{F}_K^* = \{(F, F') \in \mathcal{F} \times \mathcal{F} : F \cap K \neq \varnothing\}$, $\mathcal{F}_{K'}^* = \{(F, F') \in \mathcal{F} \times \mathcal{F} : F' \cap K' \neq \varnothing\}$.

Equiped with joint capacity functionals of random closed sets (or, dually, with joint belief functions) as developped above, research efforts could focus on how to obtain joint belief functions from the knowledge of marginal belief functions, and how to derive belief functions correponding to intersections of nonempty random closed sets, for combination of evidence. This seems to be a challenging research problem!

REFERENCES

[1] Alvarez, D.A. (2009), A Monte Carlo-based method for the estimation of lower and upper probabilities of events using infinite random sets of indexable type, *Fuzzy Sets and Systems* (160), 384-401

[2] Dempster, A. (1967) Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Statist. (38) 325-339*

[3] Durante, F. and Spizzichino, F. (2010) Simi-copulas, capacities and families of level sets, *Fussy Sets and Systems (*161) 269-276

[4] Matheron, G. (1975) *Random sets and Integral Geometry*, J. Wiley

[5] Nelson, R.B. (1999) *An Introduction to Copulas,* Lecture Notes in Statistics #139, Springer-Verlag

[6] Nguyen, H.T. (1978) On random sets and belief functions, *J. Math. Anal. and Appl. (*65) 531-542

[7] Nguyen, H.T. (2006) *An Introduction to Random Sets,* Chapman and Hall/CRC

[8] Rota, G.C. (1964) On the foundations of combinatorial theory I: Theory of Mobius functions, *A.Wahrsch. (2) 340-368*

[9] Scarsini, M. (1996) Copulae of capacities on product spaces, in *Distributions with Fixed Marginals and Related Topics,* IMS Lecture Notes, Monograph Series, vol. 28, 307-318

[10] Schmelzer, B. (2012) Characterizing joint distributions of random sets by multivariate capacities, *Intern. J. Approximate Reasoning (*53) 1228-1247

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