Upper probabilities and Additivity

Massimo Marinacci
Dipartimento di Scienze Economiche
Università di Bologna
Piazza Scaravilli 2
40126 Bologna, Italy
marinacc@economia.unibo.it

Abstract

We show that a class of upper probabilities arising in many robustness models, turns out to be additive under a fairly weak condition.

Keywords. Upper probabilities, symmetric capacities

1 Introduction and Preliminaries

Symmetric and coherent Choquet capacities are a class of upper probabilities which arise in many robustness models. In this work we show that they turn out to be additive under a fairly weak condition, thus casting some doubts on the use of this class of capacities in modeling imprecise information.

Let $\mathcal{B}$ be an event σ-algebra of subsets of a space Ω. A set function $C : \mathcal{B} \to [0,1]$ is a Choquet capacity if $C(\emptyset) = 0$, and it is coherent if there exists a nonempty set M of countably additive probability measures such that $C(A) = \sup_{P \in M} P(A)$ for all $A \in \mathcal{B}$ (in the sequel, all probabilities will be countably additive).

A Choquet capacity $C : \mathcal{B} \to [0,1]$ is symmetric if there exists a non-atomic probability measure $\mu : \mathcal{B} \to [0,1]$ such that, for all $A, B \in \mathcal{B}$,

$$\mu(A) = \mu(B) \Rightarrow C(A) = C(B).$$

Kadane and Wasserman (1996, p. 1250) observe that "many robustness models used in statistics involve symmetric, coherent capacities or can be transformed into the same by a smooth, one-to-one mapping." For this reason they undertake a thorough investigation of this class of capacities, to which we refer the interested reader (Wasserman and Kadane (1992) and Kadane and Wasserman (1996)).

2 A uniqueness theorem

The main result of this paper is based on a novel uniqueness result, which should be of interest in itself. We now state the version of this result which is relevant for the present setting. A more general version can be found in Marinacci (1998).

Theorem 1 *Let P and Q be two probability measures defined on $\mathcal{B}$. Suppose that P is non-atomic. If there exists an event $A \in \mathcal{B}$ with $0 < P(A), Q(A) < 1$ and such that*

$$P(A) = P(B) \Rightarrow Q(A) = Q(B) \tag{1}$$

whenever $B \in \mathcal{B}$, then $P = Q$.

Notice that we require only the existence of just a single set $A \in \mathcal{B}$, with $0 < P(A), Q(A) < 1$, for which the condition (1) has to be satisfied. Moreover, only P is non-atomic, while no requirement is made on Q.

3 The main result

Given a capacity $C : \mathcal{B} \to [0,1]$, a natural and widely used measure of the degree of imprecision (or vagueness, ambiguity) associated with an event $A \in \mathcal{B}$ is given by the interval

$$\left[\inf_{P \in M} P(A), \sup_{P \in M} P(A)\right]$$

or, equivalently, by the interval $\left[\overline{C}(A), C(A)\right]$ (see, e.g., Walley (1991) ch. 5).[1]

If $\inf_{P \in M} P(A) = \sup_{P \in M} P(A)$ (equivalently, if $C(A) + C(A^c) = 1$), the event A does not involve any imprecision and all priors $P \in M$ agree on A.

Definition 1 *Let $C : \mathcal{B} \to [0,1]$ be a Choquet capacity. An event $A \in \mathcal{B}$ is non-trivial and unambiguous if $0 < C(A) < 1$ and $C(A) + C(A^c) = 1$.*

[1] $\overline{C}$ is the dual capacity of C, defined by $\overline{C}(A) = 1 - C(A^c)$ for all $A \in \mathcal{B}$.

Clearly, a Choquet capacity C is a probability measure if and only if all events $A \in \mathcal{B}$ are unambiguous. However, the next result shows that for symmetric and coherent Choquet capacities, the existence of a single non-trivial unambiguous event is enough to make them additive. Since excluding the existence of even a single non-trivial unambiguous event seems in general a very strong assumption, this result might cast some doubts on the use of symmetric and coherent Choquet capacities in modeling imprecision in Bayesian decision makers' beliefs.

Theorem 2 *Let $C : \mathcal{B} \to [0,1]$ be a coherent Choquet capacity, which is symmetric with respect to a non-atomic probability measure $\mu : \mathcal{B} \to [0,1]$. Then, there exists a non-trivial unambiguous event $A \in \mathcal{B}$ if and only if C is a probability measure. In particular, $C = \mu$.*

Remark. In Bayesian decision theory, several axiomatizations enlarge the state space by assuming the existence of an external random device with given probabilities (say, a coin flip or a roulette wheel). Let $m : \mathcal{R} \to [0,1]$ be a probability measure representing the random device, defined on a suitable σ-algebra $\mathcal{R}$. The product measure $\mu \otimes m : \mathcal{B} \otimes \mathcal{R} \to [0,1]$ has always non-trivial unambiguous events (for example, all the events of the form $\Omega \times A$, where $A \in \mathcal{R}$).

Proof of Theorem 2. We first notice that there exists a function $g : [0,1] \to [0,1]$ such that $C(A) = g(\mu(A))$ for all $A \in \mathcal{B}$. Let $A \in \mathcal{B}$ be a non-trivial unambiguous event, and let $B \in \mathcal{B}$ be such that $\mu(A) = \mu(B)$. Then, $g(\mu(A)) = g(\mu(B))$ and $g(\mu(A^c)) = g(\mu(B^c))$, so that $C(B) + C(B^c) = 1$. Consequently, $P(B) = C(B)$ for all $P \in M$ and, in particular, $P(A) = P(B)$ for all $P \in M$. By Theorem 1, this implies $P = \mu$ for all $P \in M$, and so $M = \{\mu\}$. Since $C(B) = \sup_{P \in M} P(B)$ for all $B \in \Sigma$, we conclude that $C = \mu$. $\Box$

4 An extension to upper probabilities

We say that a Choquet capacity $C : \mathcal{B} \to [0,1]$ is an upper probability if, for all disjoint $A, B \in \mathcal{B}$,

1. $C(A \cup B) \leq C(A) + C(B)$,
2. $\overline{C}(A \cup B) \geq \overline{C}(A) + \overline{C}(B)$.

It is easy to see that a coherent Choquet capacity is an upper probability, while the converse is false. Upper probabilities are therefore more general than coherent Choquet capacities and several papers have shown that they can be useful in modeling some non-deterministic phenomena (see, e.g., Grize and Fine (1987), Papamarcou and Fine (1986), Sadrolhefazi and Fine (1994)).

A very important set associated with an upper probability C is the (possibly empty) set $\mathcal{E}(C)$ of all the probability measures dominated by C, that is,

$$\mathcal{E}(C) = \{P : P(A) \leq C(A) \text{ for all } A \in \mathcal{B}\}.$$

For example, if C is coherent with respect to a set M of probability measures, then $M \subseteq \mathcal{E}(C)$.

Theorem 3 *Let $C : \mathcal{B} \to [0,1]$ be an upper probability, which is symmetric with respect to a non-atomic probability measure $\mu : \mathcal{B} \to [0,1]$. If there exists a single non-trivial unambiguous event $A \in \mathcal{B}$, then $\mathcal{E}(C)$ is the singleton $\{\mu\}$, i.e., $\mathcal{E}(C) = \{\mu\}$.*

Clearly, if there exists a subset $M \subseteq \mathcal{E}(C)$ such that $C(A) = \sup_{P \in M} P(A)$ for all $A \in \mathcal{B}$, then $C = \mu$. This is why Theorem 3 extends Theorem 2 to upper probabilities.

The proof of Theorem 3 rests on the following simple lemma, of some independent interest, which extends a result of Wasserman and Kadane (1992) (lemma 8 p. 1729).

Lemma 1 *Let C and μ be, respectively, a Choquet capacity and a non-atomic probability measure defined on $\mathcal{B}$. If there exists a function $g : [0,1] \to [0,1]$ such that $C(A) = g(\mu(A))$ for all $A \in \mathcal{B}$, then such a function g is non-decreasing. In particular, if $\mathcal{B}$ is the Borel σ-algebra of $[0,1]$ and λ the Lebesgue measure, then $g(x) = C([0,x])$.*

Proof: Suppose that $1 > x > y > 0$. Since μ is non-atomic there exist $A, B, E \in \mathcal{B}$ such that $\mu(A) = x$, $\mu(B) = \mu(E) = y$, $B \nsubseteq A$, and $E \subseteq A$. Hence, $C(B) = C(E) \leq C(A)$ because C is monotone. This implies that $g(x) \geq g(y)$, and so g is non-decreasing (the case $x = 1$ or $y = 0$ is trivial). The second part can be found in Wasserman and Kadane (1992). $\Box$

Proof of Theorem 3. Since C is symmetric w.r.t. μ, there exists a function $g : [0,1] \to [0,1]$ such that $C(A) = g(\mu(A))$ for all $A \in \mathcal{B}$. By Lemma 1, the function g is non-decreasing with $g(0) = 0$ and $g(1) = 1$. Moreover, since C is an upper probabilities, g is such that, for all $x, y \in [0,1]$ with $x + y \leq 1$, we have $g(x+y) \leq g(x) + g(y)$ and $g(1-x-y) \leq g(1-x) + g(1-y) - 1$. Hence, by Lemma 9 of Wasserman and Kadane (1992), $g(x) \geq x$ for all $x \in [0,1]$, and so $\mu \in \mathcal{E}(C)$. Proceeding as in the proof of Theorem 2, it can be shown that $\mathcal{E}(C) = \{\mu\}$. $\Box$

References

[1] Grize, Y.-L. and Fine, T.L. (1987). Continuous lower probability-based models for stationary processes with bounded and divergent time averages. *Ann. Probab.* 15: 783-803.

[2] Kadane, J. and Wasserman, L.A. (1996). Symmetric, coherent, Choquet capacities. *Ann. Statist.* 24: 1250-1264.

[3] Marinacci, M. (1998). A uniqueness theorem for convex-ranged measures. University of Toronto, mimeo

[4] Papamarcou, A. and Fine, T.L. (1986). A note on undominated lower probabilities. *Ann. Probab.* 14: 710-723.

[5] Sadrolhefazi, A. and Fine, T.L. (1994). Finite-dimensional distributions and tail behavior in stationary interval-valued probability models. *Ann. Statist.* **22**: 1840-1870.

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

[7] Wasserman, L.A. and Kadane, J. (1992). Symmetric upper probabilities. *Ann. Statist.* 20: 1720-1736.