DECISION THEORY WITH HIGHER ORDER BELIEFS

Aldo RUSTICHINI, *
CORE, 34 Voie du Roman Pays
Université Catholique de Louvain
B-1348 Louvain-la-Neuve, Belgium.

December 1991

Abstract

In existing models of decision under uncertainty where the decision maker may have multiple beliefs (vague beliefs) the different beliefs are not distinguished by their degree of reliability. This paper extends the model of vague beliefs in this direction: the decision maker may have multiple beliefs, but also has a complete order over them (a belief over his own beliefs, or higher order beliefs).

We characterize the complete orders over acts which can be held by a decision maker with higher order beliefs. Standard expected utility theory is a special case of the representation we present.

1 Introduction

On June 18, 1815, at about six thirty in the afternoon the French army was close to victory over the English army. If you had asked Napoleone Bonaparte an assessment of the probability of his final success, he would have probably answered along the following lines:

"A lot depends, you know, on what Grouchy and Blücher are going to do."

Emmanuel Grouchy, maréchal de France, had the task of pursuing the Prussian army; on the 18th, he was waiting for orders in the countryside near the present location of Louvain-la-Neuve, with thirty thousand men.

Gebhard Leberecht von Blücher was the Prussian general, located north-west of Grouchy, with more than one hundred thousand men. He was not waiting. Both could have joined the battle that was taking place between Mont-Saint-Jean and Waterloo.

"So there are as you see -*l'empereur* would continue- four possible scenarios: Grouchy joins the battle, and Blücher does not; Blücher does and Grouchy does not, and so on. In the first case I think we shall win almost surely; in the second I admit it is going to be hard: five to one is a lost battle. In the third scenario..."

*Long discussions with Elchanan Ben Porath have given the start to the project. This research begun while the author was visiting the Department of Economics at Tel Aviv University, July '91 ("Summer in Tel Aviv 1991") and was completed while he was visiting CORE (UCL) during the fall of '91. To both institutions I am grateful for hospitality and support.

and so on: four different first order beliefs. If you had then persisted, and asked Napoleone his assessment of the relative probabilities of the four scenarios he had just mentioned, he would have revealed to you what we call in this paper his higher order probabilities. Together with the belief that Napoleone had on the possible outcomes in the different scenarios, they give a complete description of his probabilistic assessment over the outcome space (*win, lose*).

The purpose of this paper is to provide a theory of decision making under uncertainty and beliefs over beliefs *without assuming* that the expected utilities under different beliefs are weighted linearly, as they are not, for example, by the people whose behavior is described in the Ellsberg paradox.

We refer the reader to *e.g.* Karni and Schmeidler [1989] for a discussion of the relevant literature, and move on to the analysis.

2 Preliminaries

Notation and formal structure are fairly standard. S denotes the state space. We always assume that S is a topological space. More specifically:

Assumption 1. *S is compact and Hausdorff.*

and is endowed with the Borel σ-algebra $\mathcal{B}(S)$. The assumption of a topological structure on the state space may be unusual, but we are ready to defend it. The most natural way of modelling the state space seems to be to formulate it as the set of valuations over the propositions that the decision maker can formulate in his language. This space, in a proper formulation, satisfies the condition **A1**; indeed it satisfies the stronger condition **A1'** below. This reason is, of course, further strenghthened by the technical advantages that the topological structure gives.

For reasons that will be discussed later, we sometimes consider state spaces which satisfy the stronger assumption:

Assumption 1'. *S is compact and metric.*

The space of consequences we consider is simplified. In this we follow Bewley [1986]; in particular the utility is linear in rewards or payoffs. Formally, acts are functions from states to consequences (which are rewards or payoffs); the set of acts is denoted by $F \equiv \{f : S \to I\!R\}$. For an act $f \in F$ and a state $s \in S$, $f(s)$ is the reward or utility in state s.

In this paper we shall consider special subspaces of the space of acts, namely $C(S)$, the continuous functions, and $B(S, \mathcal{B}(S))$, the space of bounded measurable real valued functions on $(S, \mathcal{B}(S))$. We recall below definition and basic properties of the space $\mathcal{B}(S)$. In both spaces for $x \in I\!R$ we denote the constant function equal to x over X by $\mathbf{1}x$.

To the reader who finds the case of continuous acts of no interest we emphasize that all the results discussed in the following extend to the case of measurable acts. This point is discussed in detail later. The only substantial price to pay is that the beliefs of the decision maker, when $F = B(S, \mathcal{B}(S))$, are finitely additive measures, as they are in Savage [1952].

For the Borel σ-field $\mathcal{B}(S)$ we define simple functions as the finite linear combination with real coefficients of characteristic functions of sets in $\mathcal{B}(S)$. The space $B(S, \mathcal{B}(S))$ is then the uniform limit of simple functions; it is a Banach space with norm $\|f\| = \sup_{s \in S} |f(s)|$. Since there is no ambiguity, we shall shorten the notation to $B(S)$. The dual of $B(S)$ is $bfa(S)$, the space of bounded finitely additive set functions.

The case $F = B(S)$ can either be treated separately or reduced to the case in which acts are continuous functions by a well known technique based on Stone's Representation Theorem. We follow this second way here to preserve homogeneity of the presentation. In this way in fact the analysis of $F = B(S)$ is reduced to the case $F = C(S)$ with the only technical complication that the space S cannot be assumed to be metric.

The representation theorem we use is:

Theorem 2.1. *(Dunford and Schwartz, [1957], IV.6.18 and 19). There exists a compact Hausdorff space S_1 and an isometric algebraic isomorphism U between the algebras $C(S_1)$ and $B(S, \mathcal{B}(S))$. U maps positive functions into positive functions. Also there exists a one to one embedding of S as a dense subset of S_1, such that each f in $B(S, \mathcal{B}(S))$ has a unique continuous extension f_1 to S_1, and $f_1 = U(f)$.*

This theorem allows us to identify acts which are Borel measurable functions on S as continuous functions on a different (larger, in the sense of the theorem) space, and correspondingly to identify the finitely additive measures on S with the countably additive measures on S_1. An additional step in the analysis would be to provide conditions which insure the beliefs over S itself to be σ-additive measures. This is not pursued in this paper.

When S is metric, we consider the space of regular, countably additive probability measures over $(S, \mathcal{B}(S))$, denoted by $\mathcal{P}(S)$.
It is always assumed to be endowed with the weak* topology. As such it is itself a compact metric space, and $(\mathcal{P}(S), \mathcal{B}(\mathcal{P}(S)))$ has a set of regular, countably additive probability measures defined on it. We denote this set by $\mathcal{P}(\mathcal{P}(S))$.

In the more general case of S_1 compact Hausdorff we have that $\mathcal{P}(S_1)$, the set of probability measures, is a compact subset of $rca(S_1)$, the space of regular countably additive measures on S_1, again endowed with the weak* topology. We remark that $\mathcal{P}(\mathcal{P}(S_1))$ is a non empty compact Hausdorff set, but non necessarily metric (see *e.g.* Heyer, [1977]).

With these preliminaries out of the way, we can now go back to our decision maker. He is assumed to hold a preference relationship over F which is assumed to be complete and transitive, and is denoted by $\succeq$. We furthermore assume that this preference relation can be represented by a real valued function v. Formally:

Assumption 2. *There exists a real valued function $v : F \rightarrow \mathbb{R}$ such that for any $x, y \in F$ $x \succeq y$ if and only if $v(x) \geq v(y)$.*

The space of acts is in both cases $F = C(S)$ or $B(S)$ a vector lattice with the order given by the pointwise inequality. We write, for $x, y \in F$, $x \geq y$ if and only if $x(s) \geq y(s)$ for every $s \in S$.

As in Bewley [1986: assumption 1.2] we assume that more utility is better, that is:

Assumption 3. *$x \geq y$ implies $x \succeq y$,*

or equivalently:

Assumption 3'. *v is an increasing function.*

We shall require a regularity condition on v. Stronger conditions will be considered later, but a standing assumption is the following:

Assumption 4. *v is a locally Lipschitz function.*

A function v is said to be locally Lipschitz at $x \in F$ if there exists a neighborhood U of x such that v is Lipschitz on U.

In the spirit of revealed preference analysis we now ask if a given ("revealed") preference order of a decision maker, as summarized by a function v, can be rationalized as thfe set of preferences induced by higher order beliefs. Formally, we say that the order $\succeq$ has a higher order belief representation given by a pair (V, ρ) if the following condition is satisfied.

$$x \succeq y \text{ if and only if } \int_{\mathcal{P}(S)} V(\mu(x))d\rho(\mu) \geq \int_{\mathcal{P}(S)} V(\mu(y))d\rho(\mu). \tag{2.1}$$

for a pair $V : I\!R \to I\!R$, and $\rho \in \mathcal{P}(\mathcal{P}(S))$. Here $\mu(x)$ denotes the integral of the function x with respect to the measure μ. From the definition of v we have that the function v must satisfy

$$v(x) = \int_{\mathcal{P}(S)} V(\mu(x))d\rho(\mu). \tag{2.2}$$

This is in fact the representation that we shall consider. Notice that the function $\mu \mapsto V(\mu(x))$ is continuous from $(\mathcal{P}(S), \text{weak}^*)$, and therefore the integral is well defined. Also note that for any $x \in I\!R$, $V(x) = v(\mathbf{1}x)$, and therefore v defines completely the function V. So in fact the problem we are considering is the one of determining the higher order belief ρ. A consequence of this trivial remark is an easy example of a preference order which does *not* admit a representation as in (2.2). If $v(x, y) = x^\alpha y^\beta$, with $\alpha + \beta = 1, \alpha > 0, \beta > 0$, then $V(x) = x$, and for no $\rho \in \mathcal{P}([0,1])$ we have (2.2). The case $\alpha + \beta < 1$ will require a different argument. A second consequence, which we record here for future reference, is that v is continuous if and only if V is continuous, and is Lipschitz if and only if V is.

3 Non-smooth Preferences

Given the relatively weak regularity condition on v given by **A4**, non-smooth analysis is a forced choice. Several different formulations of non-smooth analysis, each attempting to provide an extension of the concept of derivative, are available. We choose, for his simplicity, the one given by Clarke [1983].

For a Banach space X with dual X^*, let v be a real valued function on X, Lipschitz near $x \in X$. Then for any $h \in X$ we define the generalized directional derivative at x in the direction h as follows:

$$v^0(x;h) \equiv \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{v(y+th) - v(y)}{t} \tag{3.1}$$

The locally Lipschitz condition insures that the limit is finite for every pair (x, h). It is known (see Clarke [1983]) that (a) $v^0(x, \cdot)$ is finite, positively homogeneous, subadditive and Lipschitz with the same Lipschitz norm as v; and (b) $v^0(\cdot, \cdot)$ is upper semicontinuous.

The generalised gradient of v at x is the subset of X^* defined by

$$\partial v(x) \equiv \{z \in X^* : v^0(x,h) \geq z(h) \text{ for every } h \in X\}$$

where $z(h)$ is the value of z at h. It is known (see Clarke [1983]) that (a) $\partial v(x)$ is a non empty, convex, weak* compact subset of X; and (b) for every $h \in X$ one has

$$v^0(x;h) = \max\{z(h) : z \in \partial v(x)\} \tag{3.2}$$

The relationship between the traditional concepts of derivatives and the generalized gradient is discussed in Clarke [1983]. The main result is that the generalized gradient is indeed a generalization, namely that if v admits a Gâteaux, or Hadamard, or strict, or Frechet derivative at x, denoted $Dv(x)$, then $Dv(x) \in \partial v(x)$. The inclusion may be strict in pathological examples; in the case of strict differentiability, however, $\partial v(x) = \{Dv(x)\}$.

A specifically interesting example is the case in which v incorporates uncertainty adversion, that is v is concave and locally bounded. In this case v is automatically locally Lipschitz, $\partial v(x)$ coincides with the superdifferential of v at x, and $v^0(x;h)$ coincides with the directional derivative in the direction h.

In our case the assumption that v is increasing gives an additional information about $\partial v(x)$. Note that the lattice structure on F induces a vector lattice structure in the dual; we denote by X^{*+} the positive cone of the dual of X. Then:

Lemma 3.1. *Let X be a Banach lattice, and $v : X \to \mathbb{R}$ locally Lipschitz, increasing. Then*

$$\partial v(x) \subseteq X^{*+}$$

for any $x \in X$.

Proof. Let $z \in \partial v(x)$. We claim that for any $h \geq 0$, $z(h) \geq 0$. Suppose otherwise that there exists an $h \geq 0$ with $z(h) < 0$: since v is increasing, one has $v^0(x, -h) \leq 0$ directly from the definition, and from the two inequalities we conclude $v^0(x, -h) < z(-h)$. But $z \in \partial v(x)$ implies $v^0(x, u) \geq z(u)$ for every $u \in X$, a contradiction. ∎

One last mathematical preliminary. In order to give a simple integral representation of the derivative of v in the case of a non finite state space S we need to introduce the Gelfand integral of a function with values in the dual of a Banach space, $C(S)^*$ in our case. We need to this purpose the following:

Lemma 3.2. *Let $f : \mathcal{P}(S) \rightarrow C(S)^*$ be a function such that the real valued function $\mu \mapsto f(\mu)(g)$ is $\mathcal{B}(\mathcal{P}(S))$ measurable and ρ-integrable for any $g \in C(S)$; then there exists a $\nu \in C(S)^*$ such that*

$$\nu(g) = \int_{\mathcal{P}(S)} f(\mu)(g) d\rho(\mu) \text{ for every } g \in C(S).$$

For a proof of this lemma, see Diestel-Uhl [1977] p. 53. If we let $f : \mu \mapsto \mu V^1(\mu(x))$ for a fixed $x \in C(S)$, we have that for any $g \in C(S)$, the function of μ defined by $f(\mu)(g) \equiv \mu(g)V^1(\mu(x))$ is measurable and integrable, and therefore the element of $C(S)^*$ given by $\int_{\mathcal{P}(S)} \mu V^1(\mu(x)) d\rho(\mu) \equiv \nu$ is defined, and $\nu(g) = \int_{\mathcal{P}(S)} \mu(g) V^1(\mu(x)) d\rho(\mu)$, for every $g \in C(S)$.

We are now ready to show how the preferences, as represented by v, describe the higher order beliefs. The main conclusion is that the first moment of the higher order belief is completely determined by the value of v in the neighborhood of constant acts.

Theorem 3.3. *Let S be compact Hausdorff, and $v : C(S) \rightarrow I\!R$ be a locally Lipschitz function satisfying (2.2) above. Then*

$$\partial v(\mathbf{1}x) = \partial v(\mathbf{1}x)(S) \int_{\mathcal{P}(S)} \mu d\rho(\mu)$$

Remark. Note that there exists a unique probability measure in $\partial v(\mathbf{1}x)$, and it is given by the first order average belief $\int_{\mathcal{P}(S)} \mu d\rho(\mu)$.

Proof. We begin with two preliminary claims:

$$\partial V(x) \subset \partial v(\mathbf{1}x)(S) \text{ for every } x \in I\!R \tag{3.3}$$

$$\partial V(\mu(\cdot))(\mathbf{1}x) \subset \mu(\partial V)(\mu(\mathbf{1}x)) \text{ for every } x \in I\!R, \mu \in \mathcal{P}(S). \tag{3.4}$$

Since $V(x) = v(\mathbf{1}x)$ for any $x \in I\!R$, denoting by $h : I\!R \rightarrow F$ the linear functions $x \rightarrow \mathbf{1}x$ we have from the Chain Rule (Clarke [1983], 2.3.10) that $\partial V(x) \subset \partial v(\mathbf{1}x) \circ h$. On both sides of the inclusion we have a closed interval in the real line; applying both sides to $y \in I\!R$, we have from $\partial v(\mathbf{1}x) \circ h(y) = y\partial v(\mathbf{1}x)(S)$, which is what claim (3.3) says. The claim (3.4) follows also from the Chain Rule.

We then claim that for the case of S compact Hausdorff

$$\partial v(\mathbf{1}x) \subset \int_{\mathcal{P}(S)} \mu(\partial V)(\mu(\mathbf{1}x)) d\rho(\mu). \tag{3.5}$$

We consider first the case in which S is also metric. Here the claim (3.5) follows from (3.3), (3.4) and the formula

$$\partial \int_{\mathcal{P}(S)} V(\mu(x))d\rho(\mu) \subset \int_{\mathcal{P}(S)} \partial V(\mu(\cdot))(x)d\rho(\mu) \tag{3.6}$$

for any $x \in C(S)$, (see Clarke [1983] p. 76). It is easy to check that the conditions required for the above differentiation under the integral sign (respectively 2.7.1(i) and (ii), p. 75, and (c), p. 76) are satisfied in this case. In particular note that the condition (c) requires $\mathcal{P}(S)$ be a separable metric space: a requirement which is obviously satisfied when S is compact and metric.

We may now consider the case of S compact Hausdorff, and we claim that (3.6) holds in this case too. As in Clarke, from the definition of generalized directional derivative and Fatou's lemma we obtain:

$$\int_{\mathcal{P}(S)} V(\mu(\cdot))^0(x,h)d\rho(\mu) \geq \left(\int_{\mathcal{P}(S)} V(\mu(\cdot))d\rho(\mu)\right)^0 (x,h) \geq z(h) \tag{3.7}$$

for any $z \in \partial \int_{\mathcal{P}(S)} V(\mu(\cdot))(x)d\rho(\mu)$ and $h \in C(S)$. We note first that the correspondence ∂V has a simple form. Recall that

$$V(\mu(\cdot))^0(x,h) \equiv \max\{\partial V(\mu(\cdot))(x)(h)\}.$$

Since h is fixed, we may partition $\mathcal{P}(S) = \mathcal{P}(S)^+ \cup \mathcal{P}(S)^-$, where $\mathcal{P}(S)^+ \equiv \{\mu : \mu(h) \geq 0\}$, $\mathcal{P}(S)^- \equiv \{\mu : \mu(h) < 0\}$, two Borel measurable subsets. Then

$$\max\{\partial V(\mu(\cdot))(x)(h)\} = \max\{\mu(h)(\partial V)(\mu(x))\},$$

and this last is equal to $(\max(\partial V)(\mu(x)))\mu(h)$ if $\mu \in \mathcal{P}(S)^+$, and $(\min(\partial V)(\mu(x)))\mu(h)$ if $\mu \in \mathcal{P}(S)^-$. We now consider more closely the set $\partial V(\mu(x))$. Since V is defined over $I\!R$, by positive homogeneity of $V^0(y,\cdot)$ for every $y \in I\!R$ the two points $V^0(y,\pm 1)$ completely describe the function $V^0(y,\cdot)$; also V is increasing, so $V^0(y,-1) \leq 0 \leq V^0(y,1)$, and therefore clearly

$$\partial V(y) = \{[-V^0(y,-1), V^0(y,1)]\}.$$

We now proceed with the proof. We want to prove the inclusion

$$\partial \int_{\mathcal{P}(S)} V(\mu(\cdot))(x)d\rho(\mu) \subset \int_{\mathcal{P}(S)} \partial V(\mu(\cdot))(x)d\rho(\mu).$$

The integral on the right has to be interpreted as the Gelfand integral, and that the integrand is equal to $\mu(\partial V)(\mu(x))$. Since the integrand is a correspondence, let us be more specific on the meaning of this integral. Let $\mathcal{S}$ be defined by

$$\mathcal{S} = \{f : \ \mathcal{P}(S) \to I\!R, f(\mu) \in (\partial V)(\mu(x))\rho - a.e.\ \mu, \\ f \text{ is } \mathcal{B}(\mathcal{P}(S))\text{-measurable and integrable}\}.$$

This set is not empty (the function $\mu \mapsto V(\mu(x),1)^0$ is an element in the set) and convex. With the topology $\sigma(L_\infty, L_1)$ on the real valued $L_\infty(\rho)$ functions, this set is also compact, because it is the set of measurable selections out of a real, convex, compact valued correspondence (see Castaing and Valadier [1977]).

For any $f \in \mathcal{S}$, consider now the function $F : \mu \mapsto \mu f(\mu)$, a weakly measurable function. We may now define $\int_{\mathcal{P}(S)} \partial V(\mu(\cdot))(x) d\rho(\mu)$ as the set of Gelfand integrals of such functions, and find it convenient to write $G(f)$ the Gelfand integral of F, where $F(\mu) = \mu f(\mu)$. Then we let:

$$\int_{\mathcal{P}(S)} \partial V(\mu(\cdot))(x) d\rho(\mu) \equiv \{z \in C(S)^* : z = G(f), \text{ for some } f \in \mathcal{S}\}.$$

Consider now a net $\{f^\nu\}$, with $f^\nu \to f^0$ in $\sigma(L_\infty, L^1)$, and let $F^\nu = \mu f^\nu(\mu)$. Then $G(f^\nu) \to G(f^0)$ in the weak* topology of $C(S)^*$. In fact for any $h \in C(S)$,

$$\begin{aligned} G(f^\nu)(h) &= \left(\int_{\mathcal{P}(S)} \mu f^\nu(\mu) d\rho(\mu)\right)(h) \\ &= \int_{\mathcal{P}(S)} \mu(h) f^\nu(\mu) d\rho(\mu) \to \int_{\mathcal{P}(S)} \mu(h) f^0(\mu) d\rho(\mu) \\ &= G(f^0) \end{aligned}$$

because $\mu \mapsto \mu(h)$ is in $L_1(\rho)$, for every $h \in C(S)$.
As a continuous image of a compact set, $\int_{\mathcal{P}(S)} \partial V(\mu(\cdot))(x) d\rho(\mu)$ is weak* compact. Let now $z \in \partial \int_{\mathcal{P}(S)} V(\mu(\cdot))(x) d\rho(\mu)$. We have seen that

$$\sigma(h; \partial \int_{\mathcal{P}(S)} V(\mu(\cdot))(x) d\rho(\mu)) \leq \sigma(h; \int_{\mathcal{P}(S)} \partial V(\mu(\cdot))(x) d\rho(\mu)),$$

for every $h \in C(S)$; here $\sigma(h; A)$ is the support function of the set $A \subset C(S)^*$ at h. This inequality is (3.7) above, rewritten. By Hörmander's theorem (Hörmander [1954]), we have the inclusion. An easy argument by contradiction shows that the two sides are actually equal. ∎

4 Differentiable Preferences

When V is differentiable of order k, the function v allows us to identify higher order moments of the measure ρ: they are immediately determined by the higher order derivatives of v. We first define precisely this last notion. In section 3 the use of the Gelfand integral has allowed us to define the first order derivative of v, by identifying the value of the derivative at a point $x \in C(S)$ with an element of $C(S)^*$ which had a simple integral representation. The argument used to prove Lemma 3.2 above can also be used to provide the same representation for higher order derivatives. More precisely, for any k such that V^k exists and is continuous there exists a multilinear continuous functional $L_k(x, \cdot) : C(S)^k \to I\!R$ such that

$$L_k(x, h_1, \cdots, h_k) = \int_{\mathcal{P}(S)} \prod_{j=1}^{k} \mu(h_j) V^k(\mu(x)) d\rho(\mu) \text{ for every } (h_1, \cdots, h_k) \in C(S)^k.$$

Then we have

Theorem 4.1. *Let V be a C^r function, and $\rho \in \mathcal{P}(\mathcal{P}(S))$. Then $v : C(S) \to I\!R$ defined as in (2.2) is a C^r-function, and*

$$\partial^k v(x)(h_1, \cdots, h_k) = \int_{\mathcal{P}(S)} \prod_{j=1}^{k} \mu(h_j) V^k(\mu(x)) d\rho(\mu)$$

for every $k \leq r$, $(h_1, \cdots, h_k) \in C(S)^k$, *and* $x \in C(S)$.

The proof is left to the reader.

In particular, if V is smooth, then the measure ρ is determined, as we see from the following:

Theorem 4.2. *Let* V *be a* C^∞*-function, and* $V^k(x) \neq 0$ *for every* k *and for some* x. *Then the measure* ρ *is uniquely determined by the sequence of derivatives of* v *at* $\mathbf{1}x$.

Proof. Consider the Fourier transform of ρ, $\hat{\rho} : C(S) \to \mathbb{C}$, where $\mathbb{C}$ is the field of complex numbers, given by

$$\hat{\rho}(x) = \int_{\mathcal{P}(S)} \exp(\mathrm{i}\mu(\mathrm{x}))\mathrm{d}\rho(\mu).$$

We first claim that if $\hat{\rho}_1(x) = \hat{\rho}_2(x)$ for every $x \in C(S)$, then $\rho_1 = \rho_2$ on $C(\mathcal{P}(S))$. Note that the equality of the transforms $\hat{\rho}_1$ and $\hat{\rho}_2$ holds only on $C(S)$, a subspace of $C(\mathcal{P}(S))$, but the algebra of functions $\{F_x : \mathcal{P}(S) \to \mathbb{C}, x \in C(S)\}$, where $F_x(\mu) = \exp(\mathrm{i}\mu(\mathrm{x}))$ satisfies the conditions of the Stone-Weierstrass Theorem.

So if $\rho_1(g) = \rho_2(g)$ for every g in this algebra, then the equality holds for $h \in C(\mathcal{P}(S))$. Now if $\rho_1(h) = \rho_2(h)$ for every $h \in C(\mathcal{P}(S))$, then $\rho_1 = \rho_2$ on $\mathcal{B}(\mathcal{P}(S))$, and we conclude the proof of the claim.

The theorem now follows from the fact that

$$\hat{\rho}(x) = \sum_{j=0}^{\infty} \frac{i^j}{j!} \int_{\mathcal{P}(S)} \mu(x)^j d\rho(\mu)$$

from the dominated convergence theorem, and this last term is in turn equal to

$$\sum_{j=0}^{\infty} \frac{i^j}{j!} \frac{\partial^j v(\mathbf{1}x)}{V^j(x)}.$$

■

5 Finite State Space

We now concentrate on the finite state space case. Let $S = \{s_1, \cdots, s_{m+1}\}$; the set $\mathcal{P}(S)$ is now Δ_n, the n-dimensional simplex; a generic element $p \in \Delta_n$ is $p = (p_1, \cdots, p_n)$; also let $e = (1, \cdots, 1) \in \mathbb{R}^n$. For a $\beta \in \mathbb{N}^{n+1}$, we shall find it useful to denote $\beta = (\gamma, i)$, with $\gamma \in \mathbb{N}^n$, $i \in \mathbb{N}$. Also $|\beta| = \sum_{i=1}^{n+1} \beta_i$.

Note that a given distribution ρ with support on Δ_n induces the sequence of moments $\{M_\rho^\beta\}_{\beta \in \mathbb{N}^{n+1}}$, defined by

$$M_\rho^\beta = \int_{\Delta_n} p_1^{\beta_1} \cdots p_n^{\beta_n} (1 - p \cdot e)^{\beta_{n+1}} d\rho(p). \tag{5.1}$$

For a function $v : \mathbb{R}^{n+1} \to \mathbb{R}$ we denote, whenever it exists, the β-derivative as

$$\partial^\beta v(x) = \left(\frac{\partial}{\partial x_1}\right)^{\beta_1} \cdots \left(\frac{\partial}{\partial x_{n+1}}\right)^{\beta_{n+1}} v(x_1, \cdots, x_{n+1}).$$

Also for a pair $(p, \gamma) \in \mathbb{R}^n \times \mathbb{N}^n$ we write $p^\gamma = \prod_{i=1}^n p_i^{\gamma_i}$.

A utility function which satisfies a representation as (2.2) also satisfies a necessary condition :

Theorem 5.1. *If $v : \mathbb{R}^{n+1} \to \mathbb{R}$ has a representation as in (2.2), and $V^k(x) \neq 0$ for every k and for some x, then the ratios*

$$R(\beta, x) = \frac{\partial^\beta v(x, x, \cdots, x)}{V^{|\beta|}(x)} \tag{5.2}$$

are independent of x, and correspond to the moments M_ρ^β of a distribution ρ over Δ_n.

The proof is immediate.

We now consider the converse problem. Suppose that the sequence of ratios $R(\beta, x)$ is a sequence M^β independent of x. We want to find conditions on this sequence which insure the existence of a measure ρ with support on Δ_n such that the representation (2.2) is satisfied. The first step in this direction is to determine conditions for the existence of a measure ρ on Δ_n such that

$$M^\beta = M^{(\gamma, i)} = \int_{\Delta_n} p^\gamma (1 - p \cdot e)^i d\rho(p). \tag{5.3}$$

The solution to this problem is a special case of the classical problem of moments. We recall briefly here the main points of this topic.

Let $\{m^\gamma\}_{\gamma \in \mathbb{N}^n}$ be an infinite sequence of real numbers, K a closed set in $\mathbb{R}^n$. An n-dimensional distribution ρ whose spectrum is contained in K, and is a solution of $m^\gamma = \int p^\gamma d\rho(p)$ for every $\gamma \in \mathbb{N}^n$ is called a solution of the $(K, \{m^\gamma\}_{\gamma \in \mathbb{N}^n})$-moment problem. A necessary and sufficient condition for a solution to exist is given in Shohat and Tamarkin [1943], and we recall it here since we are going to use it.

To a sequence $\{a(\gamma)\}_{\gamma \in \mathbb{N}^n}$ of real coefficients, all but a finite number of which are zero, we associate the polynomial $Q(p) = \sum_\gamma a(\gamma) p^\gamma$, and we denote by $\mathcal{Q}$ the set of all such polynomials. Now a sequence $\{m^\gamma\}$ defines a linear functional $\mathcal{M}$ on $\mathcal{Q}$ by

$$\mathcal{M}(Q) = \sum_{\gamma \in \mathbb{N}^n} a(\gamma) m^\gamma. \tag{5.4}$$

Note that if the M^γ's are given as in (5.1) above, then $\mathcal{M}(Q) = \int_{\Delta_n} Q(p) d\rho(p)$. The linear functional characterizes the solution to the moment problem :

Theorem 5.2. *(Shohat-Tamarkin). A necessary and sufficient condition for the $(K, \{m^\gamma\}_{\gamma \in \mathbb{N}^n})$-moment problem to have a solution is that the functional $\mathcal{M}$ is non negative over K, that is the following condition is satisfied:*

$$\text{if } Q \geq 0 \text{ on } K, \text{ then } \mathcal{M}(Q) \geq 0.$$

To give a useful criterion to varify the condition above we first reduce the set of polynomials, for which one has to check if $\mathcal{M}$ is positive, to a smaller subset. We consider $K = \Delta_n$; if $n = 1$, then the moment problem is the Hausforff problem. Define the subset $\mathcal{R}$ of $\mathcal{Q}$ as the set of polynomials that can be written as follows:

$$\mathcal{R} = \{R \in \mathcal{Q} : R(p) = \sum_{\gamma \in \mathbb{N}^n, i \in \mathbb{N}} b(\gamma, i) p^\gamma (1 - p \cdot e)^i, b(\gamma, i) \geq 0 \text{ for every } (\gamma, i)\}. \tag{5.5}$$

Clearly polynomials in $\mathcal{R}$ are non negative on Δ_n. Using the strong law of large numbers one can prove that this set is dense in the set of continuous functions positive over Δ_n. But since the sequence of numbers $\{m^\gamma\}$ is arbitrary, a stronger approximation result is needed. This is our next step.

For any $f : \Delta_n \to \mathbb{R}$ we define the m-th Bernstein polynomial of f by :

$$B_m(f)(p) = \sum_{\{|\xi| \leq m\}} f\left(\frac{\xi}{m}\right) \begin{pmatrix} m \\ \xi \end{pmatrix} p^\xi (1 - p \cdot e)^{m - |\xi|} \tag{5.6}$$

where for $(\xi, m) \in \mathbb{N}^n \times N$, we denote $\frac{\xi}{m} = \left(\frac{\xi_i}{m}\right)_{i=1}^n$, and

$$\begin{pmatrix} m \\ \xi \end{pmatrix} = \frac{m!}{(m - |\xi|)! \xi!}.$$

It is immediate that B_m maps $\mathcal{Q}$ into $\mathcal{R}$, and that it is linear on $\mathcal{Q}$. It is therefore convenient to analyze first the action of B_m on the polynomials $P_\gamma(p) = p^\gamma$. We first note that if we denote for $(\xi, k) \in \mathbb{N}^n \times \mathbb{N}^n$, $\xi_k = \xi_1(\xi_1 - 1) \cdots (\xi_1 - k_1) \cdots \xi_m \cdots (\xi_m - k_m)$ then the equality

$$\sum_{\{|\xi| \leq m, \xi \geq k - e\}} \begin{pmatrix} m \\ \xi \end{pmatrix} p^\xi (1 - p \cdot e)^{m - |\xi|} \frac{\xi_{k-e}}{m_{|k|-1}} = p^k \tag{5.7}$$

holds. This follows from the multinomial theorem. We can rewrite each factor $\left(\frac{\xi}{m}\right)^\gamma$ in $B_m(P_\gamma)$ in a way which allows us to use (5.7) :

Lemma 5.3. *For every $(\xi, \gamma, m) \in \mathbb{N}^n \times \mathbb{N}^n \times \mathbb{N}$, $m > |\gamma|$,*

$$\left(\frac{\xi}{m}\right)^\gamma = \sum_{\gamma' \leq \gamma} c(m, \gamma, \gamma') \frac{\xi_{\gamma' - e}}{m_{|\gamma'|-1}}$$

where $c(m,\gamma,\gamma') = 0\left(\frac{1}{m}\right)$ *if* $\gamma' < \gamma$, *and* $c(m,\gamma,\gamma) = 0(1)$.

Proof. Here $\gamma' < \gamma$ means $\gamma' \leq \gamma$ componentwise, and different in at least one component. Then write $\xi^\gamma = \sum_{\gamma'\leq\gamma} c(\gamma,\gamma')\xi_{\gamma'-e}$, with $c(\gamma,\gamma')$ a fixed finite set of coefficients, and then let $c(m,\gamma,\gamma') \equiv c(\gamma,\gamma')\frac{m_{|\gamma'|-1}}{m^{|\gamma|}}$. Note that $m_{|\gamma'|-1} \leq m^{|\gamma'|}$, and so $c(m,\gamma,\gamma') \leq c(\gamma,\gamma')m^{|\gamma'|-|\gamma|} = O\left(\frac{1}{m}\right)$ if $|\gamma'| < |\gamma|$ as claimed. ∎

Now from Lemma 5.3 and 5.7 above we conclude

$$B_m(P_\gamma)(p) = p^\gamma + \frac{1}{m}\sum_{\gamma'\leq\gamma} c'(m,\gamma,\gamma')p^{\gamma'}, c(m,\gamma,\gamma') = 0(1) \text{ in } m,$$

and so for $Q(p) = \sum_{\gamma\in I\!N^n} a(\gamma)p^\gamma$, $m > |\gamma|$,

$$B_m(Q)(p) = Q(p) + \frac{1}{m}\sum_{\gamma\in I\!N^n,\gamma'\leq\gamma} c'(m,\gamma,\gamma')p^{\gamma'}. \tag{5.8}$$

We can now conclude that it is enough to check that the functional $\mathcal{M}$ is positive on $\mathcal{R}$:

Theorem 5.4. *For the* $(\Delta_n, \{M^{(\gamma,0)}\}_{\gamma\in I\!N^n})$*-moment problem to have a solution* ρ *it is necessary and sufficient that if* $R \in \mathcal{R}$ *then* $\mathcal{M}(R) \geq 0$.

Proof. For the sufficiency it is enough to note that $\lim_{m\to+\infty} \mathcal{M}(B_m(Q)) = \mathcal{M}(Q)$ because of (5.8); and $\mathcal{M}(B_m(Q)) \geq 0$ for every m because B_m maps $\mathcal{Q}$ into $\mathcal{R}$. ∎

This theorem provides the answer to the original question of finding conditions on the sequence $\{M^\beta\}$ that make it the sequence of moments of a measure ρ over Δ_n. Clearly these conditions have to be satisfied by the subsequence $\{M^{(\gamma,0)}\}$; and we shall see soon that this subsequence contains all the information about ρ. But first we use Theorem 5.4 to express the condition for the existence of a solution to the moment problem in terms of a set of inequalities involving $\{M^{(\gamma,0)}\}$:

Corollary 5.5. *The* $(\Delta_n, \{M^{(\gamma,0)}\})$*-moment problem has a solution if and only if for every* $(\gamma', i) \in I\!N^n \times I\!N$

$$\sum_{j=0}^{i}\sum_{|\gamma|=j}(-1)^j\binom{i}{\gamma}M^{(\gamma+\gamma',0)} \geq 0.$$

Proof. From Theorem 5.4 it is enough to check that the linear functional induced by $M^{(\gamma,0)}$ as in (5.4) is positive on $\mathcal{R}$, and since the polynomials in $\mathcal{R}$ have positive coefficients for terms of the form $p^{\gamma'}(1-p\cdot e)^i$, it is enough to check positivity on them. But

$$p^{\gamma'}(1-p\cdot e)^i = \sum_{j=0}^{i}\sum_{|\gamma|=j}(-1)^j\binom{i}{\gamma}p^{\gamma+\gamma'}.$$

∎

We summarize the previous discussion: for a measure ρ on Δ_n to exist which satisfies $M^\beta = \int_{\Delta_n} p^\gamma (1 - p \cdot e)^i d\rho(p)$ for every $\beta = (\gamma, i) \in I\!N^{n+1}$ it is necessary and sufficient that (1) the sequence $\{M^{(\gamma,0)}\}_{\gamma \in I\!N^n}$ satisfies the condition in Corollary 5.5, and (2) that for the solution ρ determined in (1), $M^\beta = M^\beta_\rho$ for every $\beta \in I\!N^{n+1}$.

5.1 An example. (Cobb-Douglas utility).

For $v(x_1, \cdots, x_n) = \prod_{j=1}^n x_j^{\alpha_j}$, $\alpha_j \geq 0$, $\sum_{j=1}^n \alpha_j < 1$, the sequence of ratios

$$R(\beta, x) = \frac{\prod_{j=1}^n \alpha_j(\alpha_j - 1) \cdots (\alpha_j - \beta_j + 1)}{|\alpha|(|\alpha| - 1) \cdots (|\alpha| - |\beta| + 1)}$$

is a constant independent of x. A direct argument in this case shows that no representation as in (2.2) exists. For simplicity if $n = 1$ then $x_1^{\alpha_1} x_2^{\alpha_2} = \int_{[0,1]} (x_1 p + x_2(1-p))^{\alpha_1+\alpha_2} d\rho(p)$ would imply, on the line $x_2 = t x_1$, $t > 0$, $t^{\alpha_2} = \int_{[0,1]} (p + t(1-p))^{\alpha_1+\alpha_2} d\rho(p)$, and therefore $0 = \int_{[0,1]} p^{\alpha_1+\alpha_2} d\rho(p)$ and similarly $0 = \int_{[0,1]} (1-p)^{\alpha_1+\alpha_2} d\rho(p)$, a contradiction.

5.2 Uniqueness of Higher Order Beliefs

In the case we are analyzing (S finite, v a C^∞ function, $V^k(x) \neq 0$ for every k and some x) we also conclude that the higher order belief is unique. A moment problem is said to be determined if the difference among the solutions is constant. Since the measure ρ is a probability measure, a determined problem has a unique solution. From Shohat-Tamarkin ([1943], Corollary 11.3) if a moment problem has a solution with the support in a bounded set, then the problem is determined, and therefore our problem has a unique solution.

Let us remark an interesting consequence of the uniqueness result. Because of the non linearity introduced by the function V, the support of ρ is not in general a convex set. As an example of this possibility, in a two states case with $v(x, y) = \sum_{i=1}^{2} \theta_i (x p_i + y(1 - p_i))^\alpha$, $\alpha \in (0, 1)$, the uniqueness of the solution to the moment problem, with $M^{(j,k)} = \sum_{i=1}^{2} \theta_i m_i^j (1 - m_i)^k$, proves that $\mu_i = m_i \delta_{s_1} + (1 - m_i)\delta_{s_2}$ $i = 1, 2$, are the two only beliefs in the support of ρ with weights θ_i.

References

Bewley, T. 1986. *Knightian decision theory, Part I* Cowles Foundation DP n. 807.

Castaing, C. and Valadier, M. 1977. *Convex analysis and measurable multifunctions*, Lecture Notes in Mathematics n. 580, New York Berlin: Springer Verlag.

Clarke, F.H. 1983. *Optimization and non-smooth analysis.* Wiley-Interscience, Canadian M.S. Series of Monographs and Advanced Texts.

Diestel, J. and Uhl, J.J. Jr 1977. *Vector measures.* Providence, Rhode Island: American Mathematical Society.

Dunford, N. and Schwartz, J.T. 1957 *Linear operators, Part I.* New York: Wiley Interscience.

Heyer, H. 1977. *Probability measures on locally compact groups.* New York Berlin: Springer-Verlag.

Hörmander, L. 1954. *Sur la Function d'appui des ensembles convexes dans un espace localement convexe.* Arkiv for Math. 3, 181-186.

Karni, E. and Schmeidler, D. 1989. *Utility theory with uncertainty.* To appear in *Handbook of Mathematical Economics*, volume 4; W. Hildenbrand and H. Sonnenschein eds.

Savage, L.J. 1952. *The foundation of Statistics*, New York: Wiley. Second edition 1972, New York: Dover.

Shohat, J.A. and Tamarkin, J. D. 1943. *The problem of Moments.* Providence, Rhode Island: American Mathematical Society. Mathematical Surveys, number 1.