Updating Non-Additive Probabilities – A Geometric Approach*

Ehud Lehrer†‡

May 19, 2004

Abstract

A geometric approach, analogous to the approach used in the additive case, is proposed to determine the conditional expectation with non-additive probabilities. The conditional expectation is then applied for (i) updating the probability when new information becomes available; and (ii) defining the notion of independence of non-additive probabilities and Nash equilibrium.

---

*The first draft of this paper, dated May 1996, was distributed under the title "An Axiomatic Approach to the Conditional Expectation with Non-Additive Probabilities".

†School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel; e-mail: lehrer@post.tau.ac.il; web site: www.math.tau.ac.il/∼lehrer

‡Conversations with David Schmeidler, to whom I am indebted in many respects, were very helpful. My interest in this topic was inspired by Michéle Cohen. I am also grateful to Alain Chateauneuf, Eran Shmaya and Peter P. Wakker for their comments.

1 Introduction

The issue of updating non-additive probabilities (Schmeidler (1989)) has been given extensive attention. Several theories have been proposed for the conditional probability in the non-additive case (see Dempster (1967,1968), Shafer (1976), Smets (1986), Gilboa (1989), Chateauneuf and Jaffray (1989), Fagin and Halpern (1989), Halpern and Tuttle (1989), Jaffray (1990), Gilboa and Schmeidler (1993), Denneberg (1994) and Sarin and Wakker (1998)). Most suggest that the probability of an event B conditioned on an event A depends not only on the probabilities of A, B and $A \cap B$, as in the traditional Bayes formula, but also on the probabilities of other events, such as $\overline{A} \cap B$ and $(A \cap B) \cup \overline{A}$. Once the conditional probability given A is defined, say, $P(\cdot|A)$, one may define the conditional expectation of a function X (e.g., a state-of-nature-dependent payoff, derived from a certain action – an act in the terminology of Savage (1954)), given the event A, by simply integrating the restriction of X over A with respect to the conditional probability $P(\cdot|A)$.

This method of calculating the conditional expectation is conceptually inconsistent for the following reason. While the conditional probability of B with respect to A depends on the behavior of B outside of A, the conditional expectation of X, given A, depends only on the behavior of X over A. Thus, two functions may be significantly different on the complement of A, and yet, as long as they coincide on A, their conditional expectations are equal.

A similar method of calculating the conditional expectation is to restrict the probability and the function to the conditioned event and to consider only the restricted items. More precisely, the conditional expectation is defined as the Choquet integral (see Choquet (1953-1954)) of the restricted function with respect to the normalized restricted probability. This method implies that the derived conditional probability of an event B, given A, depends only on the probability of $A \cap B$ and of A. It may also imply that the conditional expectation of a function X on A is equal to its conditional expectation over $\overline{A}$ and yet, both differ from the Choquet integral of X.

In this paper, we present a geometric approach, inspired by the theory of additive probabilities, which suggests a theory of conditional expectation that does not pass through the conditional probabilities. Rather, the conditional probability is a by-product.

The conditional expectation of a function X, given a field of events, say, $\mathcal{F}$, will be defined as the closest (in some formal sense) function, which is $\mathcal{F}$-measurable. This represents a conservative attitude: the conditional expectation of a function X is another function that first, is compatible with the information (modeled by a field of events) and second, is the closest to the original one.

Here, in the case of non-additive probabilities, we adopt the same approach. There are several ways, though, to borrow this idea; all are equivalent in the additive case. It turns out that only a few, and the one presented in Section 6 among them, maintain the following two desirable properties:

(a) If X is $\mathcal{F}$-measurable, then the expectation conditioned on $\mathcal{F}$ is X.

(b) The expectation of a function conditioned on the trivial field is equal to its (Choquet) integral.

Some points are worth noticing. The conditional probability of an event B with respect to (w.r.t.) another event A, is the expectation of $\mathbb{1}_B$ (the characteristic function of B) conditioned on the field generated by A (which consists of ϕ, Ω, A and $\overline{A}$). However, if the complement of A is split differently (to more than just $\overline{A}$), then the conditional expectation on A itself is typically different. In other words, the conditional probability of B, given A, depends on the partition of $\overline{A}$. This observation calls for a re-examination of the concept of probability in the non-additive case.

The original probability and the conditional one must be of the same nature. In the case of additive probabilities no problem arises: the conditionals do not depend on the entire partition of the space. The conditional probability depends on the conditioned event only, and therefore it is a numeric probability, like the original one. In the non-additive case, by contrast, the conditional probability depends on the whole partition and is not real-valued. Rather, it is a function: the conditional probability of B, given the partition $\mathcal{P}$, is a $\mathcal{P}$-measurable function. That is, the conditional probability is typically vector-valued. The unconditional probability must, therefore, be considered as a special case of the conditional probability, given the **trivial field**, and hence, real-valued.

All of the above suggest that the probability must be defined as a vector whose dimension depends on the partition under consideration.

In case the conditional expectation is not computed directly, but rather through the conditional probability, an undesirable phenomenon occurs: the conditional expectation of $\mathbb{1}_B$, given A, does not coincide with the probability of B given A. When computed directly, as hereby proposed, this does not happen. Moreover, due to continuity, if a function is close to $\mathbb{1}_B$, then its conditional expectation, given A, is an approximation to the probability of B, given A.

Under some existing updating schemes it may turn out that the conditional probabilities of B, given A, and of B, given $\overline{A}$, are both less than some constant, and yet, the probability of B is greater than this constant. Under the updating scheme proposed here this cannot occur. This feature extends to the conditional expectation. The fact that the conditional expectation of a function is uniformly greater than a certain constant implies that the integral of this function is greater than the same constant. In particular, if, given any event in the informational partition, an act is valued, say, 7, then this act is unconditionally valued 7.

This approach is used to define Nash equilibrium with non-additive probabilities. In a strategic context, Nash equilibrium involves two conditions. First, the players play independently, and thus, their play induces independent probabilities over the product of their action spaces. Second, each player plays his or her best response, given his or her choice and given other players' actions.

In case the mixed actions of the players are non-additive, the first condition calls for a definition of independence of non-additive probabilities defined on a product space. Using the geometric approach suggested here, this has a natural solution: the mixed actions of the players are independent if there is a measure over the set of all joint actions such that (i) the marginal probability over every player's actions coincides with the players' mixed action; and (ii) the players can induce nothing about other players' actions from their own. Only through conditional probability can players learn about others' actions from their own. Therefore, condition (ii) of independence can be conveyed more formally as follows. There exists a probability over the product space (typically, **not** the **product** probability) such that the probability of player i playing an action in a set B coincides with the conditional probability of B, given the partition induced by what player i knows (i.e., his or her actions). Section 7 elaborates on this subject.

The second condition of Nash equilibrium refers to incentive compatibility. It states that each player plays his or her best response to other players' actions. However, the payoff given to a player when he or she plays an action, is nothing but the conditional payoff, with respect to the independent probability (over the product space), given that action. Therefore, both conditions of Nash equilibrium require the concept of conditional expectation provided here.

2 A Motivating Example

Browsers in a car dealership were asked to fill out a short questionnaire. The information they had to provide was whether they had bought a new car during the last three years (indicated as "frequent buyer") or not (indicated as "non-frequent buyer"), and the number of years they had spent in school. Due to the limited patience of the average browser, the questionnaire was designed to elicit information in steps: First, whether she/he had purchased a new car in the last three years, then whether the number of years in school was between 0 and 12 or greater than 12, and finally to indicate the number of school years in one of the following ranges: $0-8$, $9-12$, $13-15$ and 15+.

It turned out that 1000 customers filled out the form. Some, as expected, failed to provide all the information. Others skipped the second step and indicated, for instance, that they had spent between 0 and 8 years in school, without marking the $'0-12'$ category. Still other customers checked two complementing categories, such as $13-15$ and 15+. The less educated browsers were the most impatient and frequently failed to hand in fully answered forms. The number of checks in the various categories is given in the following table.

Years in School

	0 − 12		12+	
	0 − 8	9 − 12	13 − 15	15 +
frequent-buyers	200		300	
	10	100	200	100
non-frequent buyers	310		150	
	100	200	101	102

Number of checks in each category

Aside from the information provided by the table, it is known that the total number of frequent buyers is 500, of non-frequent buyers is 490, of the category "0-12" is 510, and of "12+" is 480. Finally, categories whose intersection is not mentioned did not get any joint check.

Based on the above information one can derive a non-additive probability as follows. The probability of a category is the relative frequency of the checks in this category. The same applies to an explicitly indicated intersection of categories (e.g., non-frequent buyers who have less than or equal to 12 years in school). Intersections that are not mentioned have zero probability. As for the union of categories, the probability is the sum of the probabilities of the categories that comprise the union. For instance, the probability of the non-frequent buyers who have less than or equal to 12 years in school is $\frac{310}{1000}$, the probability of the non-frequent buyers is $\frac{490}{1000}$ and the probability of the union of those non-frequent buyers who have less than or equal to 12 years in school, and those who have more than 12 years in school is $\frac{310}{1000} + \frac{480}{1000} = \frac{790}{1000}$.

This example is aimed at convincing the reader that the probability of a "non-frequent buyer", given the event "0-12", depends on all the information available. In particular, the conditional probability when the available information is the partition "0-12" and "12+" differs from the conditional probability when the available information is the partition "0-12", "13-15" and "15+".

Denote the partition consisting of the events "0-12" and "12+" by $\mathcal{F}_1$ and the partition consisting of "0-12", "13-15" and "15+" by $\mathcal{F}_2$. For the sake of example, consider a new customer who filled out a questionnaire and indicated that she/he has spent between 0 and 12 years in school. What is the conditional probability that she/he has bought a new car in the last three years?

Suppose that the information available is $\mathcal{F}_1$. Based on it, the size of category "0-12" can be directly estimated as 510, which is the total number of checks in the category "0-12" – an event in $\mathcal{F}_1$. Alternatively, one can use the dual estimation (the complement of the complement), and obtain $1000 - 480 = 520$ (since 480 is the total number of checks in the category "12+", which is also an event in $\mathcal{F}_1$). Notice that had the table reflected the true situation, the two estimation methods would result in the same number. Due to the distorted information the estimations are different. Dempster (1967) and Shafer (1976) used the dual method, but there is no convincing reason to choose one over the other.

Now assume that the information available is wider, say, $\mathcal{F}_2$. There are two additional numbers known: 301 – the number of checks in "13-15", and 202 – the number of checks in "15+". Taking these numbers at face value, using the dual estimation, one would estimate the size of "0-12" as $1000-301-202 = 497$. Thus, the additional information results in a different estimation.

The question arises as to whether one should ignore the previously known figure of 450, or maybe the newly known numbers and if not, how should one weigh all the figures together? Furthermore, if all the numbers in the complement of "0-12" are considered why discriminate against the direct estimation? The method introduced in the sequel takes into account **all** the information available.

Hopefully, the reader is now curious enough to know how.

3 The Geometric Approach in the Additive Case

For the sake of simplicity, let us assume that the underlying probability space, Ω, is finite. Let P be an additive probability. We denote by $\mathcal{D}$ the field containing all subsets of Ω. A generic subfield of $\mathcal{D}$ will be denoted by $\mathcal{F}$. If $\mathcal{F}$ is the trivial field (containing ϕ and Ω only) it will be denoted as $\mathcal{T}$. The field that consists of ϕ, Ω, A and the complement of A, $\overline{A}$, is denoted by $\mathcal{F}_A$.

Assume that X is a random variable and let $\mathcal{F}$ be a field. It turns out that X can be written as $X = Y + X^{\perp}$, where Y is $\mathcal{F}$-measurable (i.e., Y is constant on the atoms of $\mathcal{F}$) and $X^{\perp}$ satisfies

$$(1) \qquad \int ZX^{\perp}dP = 0 \quad \text{for all } \mathcal{F}\text{-measurable variables} \quad Z\ .$$

The conditional expectation $E(X|\mathcal{F})$ is equal to Y. In other words, $X =$

$E(X|\mathcal{F})+X^{\perp}$. In the appropriate space $E(X|\mathcal{F})$ is the closest $\mathcal{F}$-measurable function to X. More precisely, denote by $\mathcal{M}(\mathcal{F})$ the set of all $\mathcal{F}$-measurable variables. Then,

$$E(X|\mathcal{F}) = \text{argmin}_{Y \in M(\mathcal{F})} \int (X-Y)^2 dP \ . \tag{2}$$

In other words, Y is the closest, w.r.t. to the ℓ_2 norm variable in $\mathcal{M}(\mathcal{F})$, to X. Stated differently, $E(X|\mathcal{F})$ is the projection of X to the subspace (of variables) $\mathcal{M}(\mathcal{F})$.

Example 1

Consider an additive probability P and an event B. Let $\mathbb{1}_B$ be the characteristic function of B. A $\mathcal{T}$-measurable function is a constant, say, α. Now, $\int(\mathbb{1}_B - \alpha)^2 dP = (1-\alpha)^2 P(B) + \alpha^2 P(\overline{B})$. The minimum of this expression is attained, by equating the derivative to zero, at $\alpha = P(B)$. Thus, $E(\mathbb{1}_B|\mathcal{T}) = P(B)$. If instead, the field is $\mathcal{F}_A$ for some event A, then an $\mathcal{F}_A$-measurable function is α on A and β on $\overline{A}$. Let Y be such a function. Thus, $\int(\mathbb{1}_B - Y)^2 dP = \alpha^2 P(A-B) + (1-\alpha)^2 P(A\cap B) + (1-\beta)^2 P(\overline{A}\cap B) + \beta^2 P(\overline{A}-B)$. The minimum of this expression is at $\alpha = \frac{P(A\cap B)}{P(A)}$ and $\beta = \frac{P(\overline{A}\cap B)}{P(\overline{A})}$. These are exactly the conditional probabilities of B, given A, and of B, given $\overline{A}$, respectively.

4 The Non-Additive Case - Preliminary Attempts

This section provides the various considerations that might guide one before defining conditional expectation. Let P be a monotonic non-additive probability. That is, $P(\emptyset) = 0$, $P(\Omega) = 1$ and if $A \subseteq B$, then $P(A) \leq P(B)$. We now use the geometric approach presented in the previous section. This approach can be interpreted in various ways. This section introduces some possible, however imperfect interpretations. The impatient reader is advised to skip it and move directly to Section **??**, where the definition proposed by this paper is provided.

The right side of (2) can be written in various ways. Here is a sample:

$$(I) \qquad \text{argmin}_{Y\in\mathcal{M}(\mathcal{F})} \int X^2 + Y^2 - 2XYdP$$

$$(II) \qquad \text{argmin}_{Y\in\mathcal{M}(\mathcal{F})} \int X^2 + Y^2 dP - 2\int XYdP$$

$$(III) \qquad \text{argmin}_{Y\in\mathcal{M}(\mathcal{F})} \int Y^2 - 2\int XYdP$$

$$(IV) \qquad \text{argmax}_{Y\in\mathcal{M}(\mathcal{F})} \int 2XY - Y^2 dP$$

$$(V) \qquad \text{argmax}_{Y\in\mathcal{M}(\mathcal{F})} \int 2XY - Y^2 - X^2 dP.$$

In the case where P is non-additive and the integral is understood as the Choquet integral, no two of these methods are equivalent.

Whatever method is adopted, it seems natural to require that the conditional expectation would satisfy the following two desirable properties:

$$\textbf{(A1)} \qquad E(X|\mathcal{F}) = X \;\text{ if }\; X \text{ is } \mathcal{F}-\text{measurable}.$$

$$\textbf{(A2)} \qquad E(X|\mathcal{T}) = \int XdP.$$

(A1) states that if X is already measurable with respect to the field $\mathcal{F}$, the expectation of X, conditional on $\mathcal{F}$, is X itself. **(A2)** states that with respect to the trivial field, that is, when no information is available, then the conditional expectation coincides with the Choquet integral of X.

Example 2

$\Omega = \{a, b\}$, $P(a) = P(b) = 0$, $P(ab) = 1$. $X(a) = 1$, $X(b) = 10$. Let $Y(a) = \alpha$ and $Y(b) = \beta$. That is, Y is $\mathcal{D}$-measurable. Then, adopting method (III), one obtains, when $\alpha < \beta$, $\int Y^2 - 2\int XY = \alpha^2 - 2\alpha$. The minimum, which is -1, is obtained when $\alpha = 1$ and $1 < \beta$. If, on the other hand, $\beta \le \alpha$ then $\int Y^2 - 2\int XY = \beta^2 - 2\min(\alpha, 10\beta)$. The minimum of -100 is achieved when $\beta = 10$ and $\alpha = 100$. Thus, the global minimum is obtained when $Y(a) = 100$ and $Y(b) = 10$. The function at which the minimum is obtained satisfies, in particular $Y \neq X$. This is not desirable, since the field considered is $\mathcal{D}$, the same field generated by X itself.

We now follow method (IV) and compute $\text{argmax}_{Y \in \mathcal{M}(\mathcal{D})} \int 2XY - Y^2 dP$. One can see that,

$$(2XY - Y^2)(\omega) \le 2 \cdot X(\omega)X(\omega) - X^2(\omega) \ , \quad \omega = a, b \ .$$

Thus, $\text{argmax}_{Y \in \mathcal{M}(\mathcal{D})} \int 2XY - Y^2 dP = X$. In other words, if method (IV) is adopted, then the conditional expectation with respect to the field generated by X is X itself.

In this example method (III) does not satisfy **(A1)** while method (IV) does.

Example 3

Let $\Omega = \{a, b\}$, $P(a) = P(b) = 0.6$, $X(a) = 0$, $X(b) = 1$. Thus, $\int X = 0.6$. Adopting the method (I), if $Y = \alpha$ then $\int (X - \alpha)^2 dP = \alpha^2 + [(1-\alpha)^2 - \alpha^2]0.6$ when $\alpha \le \frac{1}{2}$ and $\int (X - \alpha)^2 dP = (1 - \alpha)^2 + [\alpha^2 - (1 - \alpha)^2]0.6$ when $\alpha \ge \frac{1}{2}$. The minimum when $\alpha \le \frac{1}{2}$ is attained at $\alpha = \frac{1}{2}$ and when $\alpha \ge \frac{1}{2}$ it is also attained at $\alpha = \frac{1}{2}$. The global minimum is therefore achieved by $\alpha = \frac{1}{2}$. Thus, $\text{argmin}_{Y \in \mathcal{M}(\mathcal{T})} \int X^2 + Y^2 - 2XY dP \neq \int X dP$ and(A2) is not satisfied by method (I).

It turns out that none of the methods described above satisfies both **(A1)** and **(A2)**.

Another desirable property of the conditional expectation is continuity. When the conditional expectation is calculated by an additive probability, induced by the original non-additive one, there is a lack of continuity. By contrast, all reasonable methods inspired by the geometric approach imply continuity. Since $E(\cdot|\mathcal{F})$ may be a set of solutions, the full formal meaning of the following **(A5)** will be given in the next section. (The skip from **(A1)** and **(A2)** to **(A5)** is for the sake of consistency with Section 6.)

(A5) $E(\cdot|\mathcal{F})$ is continuous.

It turns out that the conditional expectation, if defined by methods (I) or (II), satisfies **(A5)**.

Methods (I) and (II) also satisfy,

(A6) If c is a positive constant, then the conditional expectation of cX is c times the conditional expectation of X.

However, due to the non-additivity of the underlying probability, these methods do not satisfy,

(A7)$^-$ If c is a constant then the conditional expectation of $c + X$ is c plus the conditional expectation of X.

5 Updating Non-Additive Probabilities – An Illustration

Any definition of the conditional expectation implies a definition of the **conditional probability** of an event given any field. Consider two events B and A. The conditional probability $P(B|A)$ is the updating of the probability of the event A when the information received is according to the field $\mathcal{F}_A$. That is, if $\omega \in A$ is realized then the information received is A. Otherwise, the information is the complement of A. Formally, the conditional probability $P(B|A)$ is defined as the value of $E(\mathbb{1}_B|\mathcal{F}_A)$ on A.

In order to illustrate the main idea, we provide here the definition of the conditional probability of an event B, given the field $\mathcal{F}_A$, induced by method (IV).

Let Y be equal to α on A and β on $\overline{A}$. The integral is equal to

$$\int 2XY - Y^2 dP = -\beta^2 + (-\alpha^2 + \beta^2)P(A \cup B) + 2\alpha P(B) + (2\beta - \beta^2 - 2\alpha + \alpha^2)P(\overline{A} \cap B), \tag{3}$$

when $\alpha \leq \beta$. It attains its maximum at the points

$$\alpha = \frac{P(B) - P(\overline{A} \cap B)}{P(A \cup B) - P(\overline{A} \cap B)} \quad \text{and} \quad \beta = \frac{P(\overline{A} \cap B)}{1 + P(\overline{A} \cap B) - P(A \cup B)}. \tag{4}$$

(4) is consistent with $\alpha \leq \beta$ if and only if $\alpha \leq P(B)$. If $P(B) < \alpha$, then the maximum of $\int 2XY - Y^2 dP$ occurs when $\beta \leq \alpha$.

Similarly to the former case, when $\beta \leq \alpha$, the maximum of $\int 2XY - Y^2 dP$ is attained at

$$\alpha = \frac{P(A \cap B)}{1 + P(A \cap B) - P(\overline{A} \cup B)} \quad \text{and} \quad \beta = \frac{P(B) - P(A \cap B)}{P(\overline{A} \cup B) - P(A \cap B)}. \tag{5}$$

(5) is consistent with $\beta \leq \alpha$ if and only if $\beta \leq P(B)$.

To summarize, there may be three types of maxima: (4), (5) or $\alpha = \beta = P(B)$. Notice that (4) and (5) may both be consistent. In this case, the solution is the maximal one.

Example 4

Let $P(B) = 0.1$, $P(B \cap A) = P(B \cap \overline{A}) = 0.05$, $P(B \cup A) = P(B \cup \overline{A}) = \frac{3}{4}$. In this case $E(\mathbb{1}_B|\mathcal{F}_A)$ is either $\frac{1}{14}$ on A and $\frac{1}{6}$ on $\overline{A}$ or vice versa: $\frac{1}{6}$ on A and $\frac{1}{14}$ on $\overline{A}$. Thus, there are two solutions that differ from each other on a set whose probability is 1.

Let us compute $\int \mathbb{1}_B - E(X|\mathcal{F}_A)dP$. It is equal to $-\frac{1}{6}\cdot 1 + \left(\frac{1}{6} - \frac{1}{14}\right)\frac{3}{4} + \left(1 - \frac{1}{6} + \frac{1}{14}\right)\frac{1}{10} + \left(1 - \frac{1}{14} - 1 + \frac{1}{6}\right).05 = 0$.

This example suggests the following:

Proposition 1 *For every two events A and B one obtains $\int \mathbb{1}_B - E(\mathbb{1}_B|\mathcal{F}_A)dP = 0$.*

Proof. Whether the solution is (4), (5) or $\alpha = \beta = P(B)$, a direct computation proves the assertion. ■

6 The Conditional Expectation

6.1 The definition and examples

The conditional expectation of the function X, given a field $\mathcal{F}$, is an $\mathcal{F}$-measurable function that satisfies some properties. We would like to define the conditional expectation of X, given the field $\mathcal{F}$, as $\text{argmin}_{Y\in\mathcal{M}(\mathcal{F})} \int X^2 + Y^2 - 2XYdP$. The problem is that according to this method **(A2)** is not always satisfied. The correction of this flaw is performed as follows.

Denote $\underline{X}(\omega) = \min_{\omega'\in\mathcal{F}(\omega)} X(\omega')$, where $\mathcal{F}(\omega)$ is the atom of $\mathcal{F}$ containing ω. Similarly denote $\overline{X}(\omega) = \max_{\omega'\in\mathcal{F}(\omega)} X(\omega')$. Let $\mathcal{N}(X,\mathcal{F})$ be the subset of those $Y \in \mathcal{M}(\mathcal{F})$ which satisfy $\int X - YdP = 0$ and $\underline{X}(\omega) \le Y(\omega) \le \overline{X}(\omega)$ for every ω.

Lemma 1 $\mathcal{N}(X,\mathcal{F})$ *is a non-empty compact set.*

Proof. Consider a field $\mathcal{F}$ whose atoms are $A_1,\ldots,A_k$. Any $Y \in \mathcal{M}(\mathcal{F})$ is a vector $(\alpha_1,\ldots,\alpha_k)$ in $\mathbb{R}^k$. That is, Y obtains the value α_i on the atom A_i, $i = 1,...,k$. Thus, $\int(X-Y)dP$ is a function of $(\alpha_1,\ldots,\alpha_k)$, say $\varphi(\alpha_1,\ldots,\alpha_k)$. By the definition of the Choquet integral, $\mathbb{R}^k$ can be split into a finite number of regions, where *a*. Each region has a non-empty interior; and

b. At each region, $\varphi(\alpha_1, \ldots, \alpha_k)$ is a summation of k polynomials of degree 1, one for each α_i. That is,

$$\varphi(\alpha_1, \ldots, \alpha_k) = \sum_{i=1}^{k} q_i(\alpha_i) \ ,$$

where $q_i(\alpha_i)$ is a polynomial of degree 1. Thus, φ is piece-wise linear. Moreover, it is continuous and monotonic. In other words, if α and β are two vectors in $I\!R^k$ and α is greater than or equal to β on every coordinate, then $\varphi(\alpha) \geq \varphi(\beta)$. Thus, $\mathcal{N}(X, \mathcal{F})$ is a closed and bounded set. Due to monotonicity and the facts $\int(X - \underline{X})dP \geq 0$ and $\int(X - \overline{X})dP \leq 0$, we conclude that $\mathcal{N}(X, \mathcal{F})$ is not empty and that there is $Y \in \mathcal{N}(X, \mathcal{F})$ such that $\underline{X}(\omega) \leq Y(\omega) \leq \overline{X}(\omega)$ for every ω. ■

Definition 1 *The* conditional expectation of X with respect to $\mathcal{F}$*, denoted* $E(X|\mathcal{F})$*, is a random variable* $Y \in \mathcal{N}(X, \mathcal{F})$ *that minimizes* $\int(X - Y)^2 dP$. *Formally,*

$$E(X|\mathcal{F}) \in \text{argmin}_{Y \in \mathcal{N}(X,\mathcal{F})} \int (X - Y)^2 dP.$$

In words, we say that Y is a conditional expectation of X given $\mathcal{F}$ if it is an $\mathcal{F}$-measurable function which minimizes the integral of the difference between X and Y squared, among the functions Y that have two properties: (i) Y is bounded between the minimum and the maximum of X in each atom of $\mathcal{F}$; and (ii) the integral of the difference between X and Y is equal to zero.

Remark 1 *Typically there is no unique solution to the problem,*

$$\min_{Y \in \mathcal{N}(X,\mathcal{F})} \int (X - Y)^2 dP.$$

We say that Y *is* $E(X|\mathcal{F})$ *if* Y *solves this minimization problem.*

6.2 Updating non-additive probabilities

The **conditional probability** of B given A, $P(B|A)$, is defined as the value of $E(\mathbb{1}_B|\mathcal{F}_A)$ on A.

Example 5

Let A be an event containing B. $E(\mathbb{1}_B|\mathcal{F}_A)$ is α on A and 0 on $\overline{A}$ (since $\mathbb{1}_B = 0$ on $\overline{A}$). The equation $\int \mathbb{1}_B - E(\mathbb{1}_B|\mathcal{F}_A)dP = 0$ has only one solution, $\alpha = \frac{P(B)}{1-P(\overline{A}\cup B)+P(B)}$, where $0/0 = 0$. That is, $\mathcal{N}(X, \mathcal{F})$ contains only one random variable. Thus,

$$P(B|A) = \frac{P(B)}{1 - P(\overline{A} \cup B) + P(B)}.$$

Note that $P(B|A)$ depends also on the probability of the event $\overline{A} \cup B$. Note also that if P is additive, $P(B|A) = \frac{P(B)}{1-P(\overline{A}\cup B)+P(B)} = \frac{P(B)}{P(A)}$.

Example 6

Let $C = B \cup \overline{A}$, where $B \subseteq A$. The conditional expectation, $E(\mathbb{1}_B|\mathcal{F}_A)$, is α on A and 1 on $\overline{A}$. Moreover, $\alpha = \frac{P(B)}{1-P(\overline{A}\cup B)+P(B)}$.

6.3 Additional examples

Example 7

Let $\Omega = \{a, b, c, d\}$. Consider the following non-additive probability: $P(a, c, d) = \frac{1}{2}$, $P(\Omega) = 1$ and 0 otherwise. Suppose that X is 0 on b and d, $X(a) = 9$ and $X(c) = 1$. The conditional expectation, $E(X|\mathcal{F}_{\{a,b\}})$, is α on $\{a, b\}$ and β on $\{c, d\}$. α and β must satisfy $\int X - E(X|\mathcal{F}_{\{a,b\}}) = 0$. Thus, (when $\alpha \geq \beta$) $-\alpha + \frac{1}{2}(\alpha - \beta) = 0$. It means that $\beta = \alpha = 0$. For the sake of completeness one should also check the case of $\alpha \leq \beta$. When doing so, one reaches the same solution.

Example 8

Consider the previous example with the change that the probability of any set containing $\{a, c\}$, unless otherwise indicated, is $\frac{1}{4}$. $\int X - E(X|\mathcal{F}_{\{a,b\}}) = 0$ implies that when $\alpha \geq \beta$, $\alpha = \frac{1}{2} - \beta$. As for minimality, when $\beta \leq \frac{1}{2}$, as required, $\int (X - E(X|\mathcal{F}_{\{a,b\}}))^2 = \beta^2 + \frac{1}{4}((1-\beta)^2 - \beta^2)$. The minimum is therefore obtained when $\beta = \frac{1}{4}$ and $\alpha = \frac{1}{4}$. Thus, the conditional expectation is a constant (and therefore necessarily coincides with the integral – see ***(A9)*** ahead.)

Example 9

Consider the previous example with the following change: the probability of any set containing $\{a\}$, unless otherwise indicated, is $\frac{1}{8}$. When $\alpha \geq \beta$, $\int X - E(X|\mathcal{F}_{\{a,b\}}) = -\alpha + \frac{1}{2}(\alpha - \beta) + \frac{1}{4} + \frac{1}{8}(9 - \alpha - 1 + \beta) = 0$. Thus, $\alpha = 2 - \frac{3}{5}\beta$. Now, if $\beta \geq \frac{1}{2}$, then $\int (X - E(X|\mathcal{F}_{\{a,b\}}))^2 = (1-\beta)^2 + \frac{1}{8}((9-\alpha)^2 - (1-\beta)^2)$. Subject to the constraint, $\alpha = 2 - \frac{3}{5}\beta$, the minimum is achieved when $\beta = \frac{1}{2}$ and $\alpha = 1.7$. However, if $\beta \leq \frac{1}{2}$, then $\int (X - E(X|\mathcal{F}_{\{a,b\}}))^2 = \beta^2 + \frac{1}{4}((1 - \beta)^2 - \beta^2) + \frac{1}{8}((9-\alpha)^2 - (1-\beta)^2)$. This expression attains its minimum within the allowable range ($0 \leq \beta \leq \frac{1}{2}$) at $\beta = 0$ and $\alpha = 2$. Out of the two solutions the latter attains the global minimum. Therefore $E(X|\mathcal{F}_{\{a,b\}})$ is 2 on $\{a, b\}$ and 0 on $\{c, d\}$.

6.4 The properties of the conditional expectation

We now list the properties of the conditional expectation as defined above. Since $E(X|\mathcal{F})$ is typically not a singleton, in what follows "$E(X|\mathcal{F})$" should be interpreted as "there is a function in $E(X|\mathcal{F})$".

(A1) If X is $\mathcal{F}$-measurable, then $E(X|\mathcal{F})$ coincides with X.

(A2) $E(X|\mathcal{T}) = \int XdP$.

(A3) $\int (X - E(X|\mathcal{F}))dP = 0$.

(A4) $E(X - E(X|\mathcal{F})|\mathcal{F}) = 0$.

(A5) $E(X|\mathcal{F})$ is continuous in X and in P. That is, for every X, P and $\varepsilon > 0$ there is $\delta > 0$ such that if $|X - X'| < \delta$ and $|P - P'| < \delta$, then $|E_P(X|\mathcal{F}) - E_{P'}(X'|\mathcal{F})| < \varepsilon$.

(A6) If c is a non-negative constant, then $E(cX|\mathcal{F}) = cE(X|\mathcal{F})$.

(A7) If Z is $\mathcal{F}$-measurable, then $E(Z + X|\mathcal{F}) = Z + E(X|\mathcal{F})$.

(A8) $E(X|\mathcal{F})$ is, on every atom of $\mathcal{F}$, between the minimum and the maximum of X.

(A9) If P is additive, then $E(X|\mathcal{F})$ coincides with the additive conditional expectation.

(A10) If $\mathcal{F}_1$ is finer than $\mathcal{F}_2$ ($\mathcal{F}_2 \subseteq \mathcal{F}_1$) and $E(X|\mathcal{F}_1)$ is $\mathcal{F}_2$-measurable, then $E(X|\mathcal{F}_1) = E(X|\mathcal{F}_2)$.

Theorem 1 *The conditional expectation always exists and satisfies* **(A1)-(A10)**.

Proof. Similarly to φ in the proof of Lemma 1 one can define $\psi(\alpha_1, \ldots, \alpha_k)$ as the value of $\int (X - Y)^2 dP$. Due to Lemma 1, $\mathcal{N}(X, \mathcal{F})$ is a non-empty and compact set. Thus, the conditional expectation is the set of the points in the non-empty set, $\mathcal{N}(X, \mathcal{F})$, at which the continuous function ψ attains its minimum. Therefore, the existence of the conditional expectation is guaranteed.

(A1), **(A2)** and **(A3)** are assured because the solutions are in $\mathcal{N}(X, \mathcal{F})$. By negation, assume that **(A4)** is incorrect. It means that 0 is not a solution for $E(X - E(X|\mathcal{F})|\mathcal{F})$. Thus, there is $Z \in \mathcal{N}(X - E(X|\mathcal{F}), \mathcal{F})$ such that $\int (X - E(X|\mathcal{F}) - Z)^2 dP$ is strictly smaller than $\int (X - E(X|\mathcal{F}))^2 dP$. Since $E(X|\mathcal{F})$ is $\mathcal{F}$-measurable, $\min_{\omega' \in \mathcal{F}(\omega)} (X - E(X|\mathcal{F}))(\omega') \leq Z(\omega) \leq \max_{\omega' \in \mathcal{F}(\omega)} (X -$

$E(X|\mathcal{F}))(\omega')$ for every ω implies that $\min_{\omega' \in \mathcal{F}(\omega)} X(\omega') \leq (Z + E(X|\mathcal{F}))(\omega) \leq \max_{\omega' \in \mathcal{F}(\omega)} X(\omega')$ for every ω. Therefore, $Z + E(X|\mathcal{F}) \in \mathcal{N}(X, \mathcal{F})$. Hence, $E(X|\mathcal{F})$ does not solve the problem $\min_{Y \in \mathcal{N}(X,\mathcal{F})} \int (X - Y)^2 dP$.

(A5) follows directly from the continuity of ψ and the fact that the Choquet integral of a non-negative (non-positive) function is smaller (greater) than or equal to zero.

(A6) holds because for every non-negative constant c, $\mathcal{N}(cX, \mathcal{F}) = c\mathcal{N}(X, \mathcal{F})$. As for **(A7)**, if Z is $\mathcal{F}$-measurable, then $\mathcal{N}(Z + X, \mathcal{F}) = Z + \mathcal{N}(X, \mathcal{F})$.

(A8) is due to the definition.

To show **(A9)**, consider an additive probability P. The $E(X|\mathcal{F})$ solves the problem $\min_{Y \in \mathcal{M}(\mathcal{F})} \int X^2 + Y^2 - 2XY dP$. It turns out that the solution is in $\mathcal{N}(X, \mathcal{F})$. Thus, minimizing over the set $\mathcal{N}(X, \mathcal{F})$ does not change the solution. Therefore, the definition adopted here coincides with the regular conditional expectation in the case of additive probabilities.

Since **(A10)** is obvious, the proof is complete. ■

Remark 2 *a. In additive probabilities what characterizes $E(X|\mathcal{F})$ is the fact that for every $A \in \mathcal{F}$*

$$\int_A E(X|\mathcal{F}) dP = \int_A X dP.$$

(A4) *is equivalent to this requirement.*

b. **(A7)** *applies also to a constant function Z.*

c. **(A3)** *and monotonicity of the Choquet integral imply that if $E(X|\mathcal{F}) \geq c$, where c is a constant, then $\int X \geq c$. Stated more generally, if Z is $\mathcal{F}$-measurable and $E(X|\mathcal{F}) \geq Z$, then $\int X - Z \geq 0$. A similar assertion obviously holds with the inverse inequality. When X is interpreted as an act, it implies that if on every atom of the informational partition, X is valued more than some constant, say, 7, then the global "worth" of this act is at least 7.*

d. The lack of additivity must fully account for the entire lack of information, as well as for ambiguity. Once these are captured, the use of the non-additive probability, for the sake of updating for instance, must be as of a probability function. In particular, whenever a certain act is equivalent to, say, 7, regardless of the prevailing event, the act itself ought to be equivalent to 7.

e. An immediate consequence of this is the following assertion. If $E(X|\mathcal{F})$ is a constant, then this constant is the Choquet integral of X. Of particular interest is the case where $X = \mathbb{1}_B$ and $\mathcal{F} = \mathcal{F}_A$. In this case the assertion means that if $P(B|A) = P(B|\overline{A})$, then both are equal to $P(B)$.

Other updating rules may result, for instance, in $P(B|A) = P(B|\overline{A}) = \frac{1}{3}$ and yet, $P(B) = \frac{1}{2}$. This phenomenon cannot happen if the conditional probability is defined by the conditional expectation as hereby suggested.

f. In my view, c. and d. are among the main reasons why one cannot restrict attention only to the conditioned event. Rather, one should treat the whole partition, including all its atoms, ***simultaneously****.*

g. **(A9)** *can be conceived as an inverse time consistency. Suppose that conditioning on a finer field results in a function which is measurable with respect to a coarser-field. This means that the additional information provided by the finer field is not valuable and the outcome is compatible with the coarser information. In this case, conditioning directly on the coarser field would result in the same function.*

7 Independence of Non-Additive Probabilities and Nash Equilibrium

Extending the notion of independence to the context of non-additive probabilities is an important issue. The main motivation for this extension is Nash equilibrium.

Let A_i be the set of actions of player i, $i = 1, ..., n$ and let $u_i : \times_{i=1}^n A_i \to I\!R$ be player i's utility function. Suppose that player i randomly chooses an action in A_i with respect to a non-additive probability P_i. This probability need not be the actual distribution according to which she randomly selects her action. The probability P_i might be the distribution that guides her choice as perceived by other players or by an outside observer.

The notion of Nash equilibrium assumes that players choose their action **independently**.

7.1 Independence of non-additive probabilities

The knowledge of each player, beyond the description of the game, consists solely of her action. Independence of P_i would therefore mean that the knowledge of her own action does not change her belief regarding the probability over other players' actions. In terms of conditional probability it means that probability over other players' actions, conditional on any subset of player i's actions, coincides with the unconditional distribution.

Let $A_{-i} = \times_{j \neq i} A_i$ and let $\mathcal{F}_i$ be the partition of A whose atoms are $\{a\} \times A_{-i}$, $a \in A_i$. The partition $\mathcal{F}_i$ represents the knowledge available to player i.

Definition 2 *A probability P over A* realizes P_i, $i = 1, ..., n$, as independent *probabilities if*

(a) *for every i and every $C \subseteq A_i$, $P(C \times A_{-i}) = P_i(C)$; and*

(b) *for every $D \subseteq A_{-i}$, $P(D \times A_i | \mathcal{F}_i) = P(D \times A_i)$.*

In order for $P_i,\ i = 1, ..., n$, to be realized as independent probabilities, there must be a probability P over the product space, A, that satisfies two conditions. Condition **(a)** states that the marginal of P over A_i coincides with P_i. Condition **(b)** states that knowing $\mathcal{F}_i$, player i does not change her belief about others' actions. In other words, the conditional probability knowing $\mathcal{F}_i$, $P_i(D|\mathcal{F}_i)$, coincides with the original probability, $P_i(D|\mathcal{T})$.

Note that in the additive case, there is a unique probability that realizes $P_i,\ i = 1, ..., n$, as independent probabilities. This is the product probability. However, in the non-additive case, typically, the product probability will **not** realize $P_i,\ i = 1, ..., n$ as independent probabilities.

At this point I have no proof of the conjecture that for any probabilities $P_i,\ i = 1, ..., n$, there is P over A that realizes them as independent. Moreover, there is no guarantee that there is a unique probability that does it.

The definition of independence of non-additive probabilities paves the way to the definition of Nash equilibrium. The next subsection take this direction one step (not more) further.

7.2 A remark on Nash equilibrium with non-additive probabilities

Nash equilibrium requires, on top of incentive compatibility conditions that players would choose their actions independently of each other. When playing the mixed action P_i,3 player i's payoff is $E(u_i|\mathcal{F}_i)$, where the expectation is taken with respect to a probability P that realized the (non-additive) mixed actions $P_i,\ i = 1, ..., n$, as independent. Note that in case there are multiple probabilities that realize $P_i,\ i = 1, ..., n$, as independent, there may be multiple expected payoffs with the same set of mixed actions.

In equilibrium, $E(u_i|\mathcal{F}_i)$ should be greater than or equal to the expected payoff guaranteed by any specific action $a \in A_i$. However, given the action $a \in A_i$, all other players still select their actions independently of each other.

Thus, The payoff associated with action $a \in A_i$ is the expectation of player i's payoff taken with respect to a probability that realizes $(P_j)_{j \neq i}$ as independent.

A precise definition of Nash equilibrium and a discussion of this notion are beyond the scope of this paper and therefore will not be presented here.

8 Final Comments and Further Problems

8.1 The Choquet integral is not essential for the definition

We have defined the conditional expectation based on the traditional Choquet integral. In fact, one can define the conditional expectation, over every partition, based on the conditional expectation on the trivial partition. Once the conditional expectation on the trivial field is defined, the general conditional expectation can be defined as the measurable function (with respect to the partition under consideration), which minimizes the distance (induced by the conditional over the trivial field), to the original function. As long as the conditional on the trivial field owns the necessary desirable properties, whether by the Choquet integral or otherwise, the general conditional expectation will possess properties **(A1)-(A10)**.

8.2 Vector-valued probabilities

We have seen that the conditional probability of an event B given A, depends on the partition of the complement of A. In fact, the conditional probability does not depend only on the conditioned event, but rather on the conditioned partition. This suggests a re-evaluation of the concept of probability. Instead of a numeric value attached to each event given another one, the approach of defining the conditional probability as a conditional expectation suggests that "generalized conditional probability" ought to associate to every partition $\mathcal{P}$ and event B a $\mathcal{P}$-measurable function. In other words, the "generalized conditional probability" must be **vector-valued**. The traditional numeric probability must be considered then as a special case: the probability conditioned on the trivial field, $\mathcal{T}$.

8.3 The value of information

Consider the value of information in the following additive case. A decision maker is informed of the atom of a partition and selects an action. In order to maximize his expected utility he chooses the action that entails the highest expected utility conditional on the prevailing atom. The value of information in this case is, thus, the incremental utility derived from knowing the partition. In other words, the value of a partition is the additional utility, compared to knowing nothing, of knowing the atom containing the realized state of nature, once it is realized. As was shown by Gilboa and Lehrer (1991), in the additive case, this value can be expressed as an additive function over the atoms. This fact has been used to axiomatize the value of information in the additive case.

Now that the conditional expectation is defined also for non-additive probabilities, one may extend the discussion about the value of information to the non-additive case. In such a case, the value of information is no longer an additive function over the atoms. This makes the analysis more challenging.

8.4 Comonotonicity and the conditional expectation

The following definition is due to Schmeidler (1989).

Definition 3 *Let X and Y be two random variables. X and Y are comonotonic if for every $\omega_1, \omega_2, \in \Omega$, $X(\omega_1) \geq X(\omega_2)$ if and only if $Y(\omega_1) \geq Y(\omega_2)$.*

Schmeidler (1989) stated an axiom that requires that if X and Y are comonotonic, then $E(X+Y) = E(X) + E(Y)$. One may require that if X and Y are comonotonic, then, $E(X+Y|\mathcal{F}) = E(X|\mathcal{F}) + E(Y|\mathcal{F})$. The conditional expectation defined above does not satisfy this property.

8.5 An axiomatic approach

It would be interesting to find a set of appealing axioms that characterize unconditional and conditional expectations with non-additive probabilities.

9 References

Chateauneuf, A. and J.-Y. Jaffray (1989), " Some Characterizations of Lower Probabilities and Other Monotone Capacities Through the Use of Moebius Inversion", *Mathematical Social Sciences*, 17, 263-283.

Choquet, G. (1953-1954), "Theory of Capacity", *Ann. l'Institute Fourier*, 5, 131-295.

Dempster, A. (1967), "Upper and Lower Probabilities Induced by Multi-Valued Mapping", *Annals of Mathematical Statistics*, 38, 325-339.

Denneberg, D. (1994), "Conditioning (Updating) Non-Additive Probabilities", *Annals of Operations Research*, 52, 21-42.

Fagin, R. and J.Y. Halpern (1991), "A New Approach to Updating Beliefs", *Uncertainty in Artificial Intelligence: Volume VI* (eds. P.P. Bonissone, M. Henrion, L.N. Kanal, and J. Lemmer), Elsevier Science Publishers, pp. 347-374. Preliminary version appeared in Conference on Uncertainty in AI, 1990, pp. 317-324.

Gilboa, I. (1989), "Additivizations of Non-additive Measures", *Econometrica* 57, 1153-1169.

Gilboa, I and E. Lehrer (1991), "The Value of Information - An Axiomatic Approach", *Journal of Mathematical Economics* 20, 443-459.

Gilboa, I. and D. Schmeidler (1993), "Updating Ambiguous Beliefs", *Journal of Economic Theory*, **59**, 33-49.

Halpern, J.Y. and M.R. Tuttle (1993), "Knowledge, Probability and Adversaries", *Journal of the ACM*, **40**, 4, pp. 917–962.

Jaffray, J.-Y. (1990), “Bayesian Updating and Belief Functions”, mimeo.

Sarin, R. and P. Wakker (1998), “Revealed Likelihood and Knightian Uncertainty”, *Journal of Risk and Uncertainty*, 16, 223-250.

Savage, L. J. (1954), *The Foundation of Statistics*, Wiley, New York.

Schmeidler, D. (1989), “Subjective Probabilities without Additivity, *Econometrica*, **57**, 571-587.

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

Smets, P. (1986), “Combining Non-distinct Evidences”, Technical report ULB-IRIDIA-86/3.