PROBABILITIES AS SIMILARITY-WEIGHTED FREQUENCIES

By

Antoine Billot, Itzhak Gilboa, Dov Samet, and David Schmeidler

November 2004

COWLES FOUNDATION DISCUSSION PAPER NO. 1492

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Probabilities as Similarity-Weighted Frequencies*

Antoine Billot,[†] Itzhak Gilboa,[‡]
Dov Samet,[§] and David Schmeidler[¶]

Revised – September 2004

Abstract

A decision maker is asked to express her beliefs by assigning probabilities to certain possible states. We focus on the relationship between her database and her beliefs. We show that, if beliefs given a union of two databases are a convex combination of beliefs given each of the databases, the belief formation process follows a simple formula: beliefs are a similarity-weighted average of the beliefs induced by each past case.

1 Introduction

A physician administers a certain treatment to her patient. She is asked to describe her prognosis by assigning probabilities to each of several possible outcomes $\Omega = \{1, ..., n\}$ of the treatment. The physician has a lot of data

*We are grateful to Larry Epstein, David Levine, Offer Lieberman, and three anonymous referees for their comments. Gilboa and Schmeidler gratefully acknowledge ISF grant no. 975/03, and Samet – ISF grnat no. 891/04 and partial financial support by the Henry Crown Institute for Business Research in Israel.

†University de Paris II, IUF, and CERAS-ENPC. billot@u-paris2.fr

‡Tel-Aviv University and Yale University. igilboa@post.tau.ac.il

§Tel-Aviv University. samet@tauex.tau.ac.il

¶Tel-Aviv University and The Ohio State University. schmeid@post.tau.ac.il

on past outcomes of the treatment, and she can readily quote the empirical frequencies of these outcomes. Yet, patients are not identical. They differ in age, gender, heart condition, and several other measurable variables that may affect the treatment outcome. Let us assume that these form a vector of real-valued variables $X = (X^1, ..., X^k)$ and that X was measured for all past cases. Thus, case j is a $(k+1)$-tuple $(x_j, \omega_j) \in \mathbb{R}^k \times \Omega$ where, $x_j \in \mathbb{R}^k$ is the value of X observed in case j, and $\omega_j \in \Omega$ is the observed outcome of the treatment in case j. The new patient is defined by the values $x_t \in \mathbb{R}^k$ of X. How should these measurements affect the probability assessment of the physician?

It makes sense to restrict attention to those past cases that had the same X values as the one at hand, and compute relative frequencies only for these data. That is, to estimate the probability of state ω by its relative frequency in the sub-database consisting of all cases j for which $x_j = x_t$. However, large as the original database may be, the sub-database of patients whose X value is identical to x_t might be quite small or even empty. Therefore, we wish to have a procedure for assessments of probabilities over Ω that makes use of data with different X values, while taking differences in these values into account.

Assume that the physician can judge which past cases are more similar to the one at hand, and which are less similar. In evaluating the probability of a state, she may assign a higher weight to more similar cases. Formally, suppose that there exists a function $s : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}_{++}$, where $s(x_t, x_j)$ measures the degree to which, in the physician's judgment, a patient whose presenting conditions are given by $x_t \in \mathbb{R}^k$ is similar to another patient whose presenting conditions are $x_j \in \mathbb{R}^k$. Given a database of past cases $((x_j, \omega_j))_j$, we suggest to assign probabilities to the possible outcomes of treatment for a new patient with conditions x_t by the formula,

$$p_t = \frac{\sum_j s(x_t, x_j) p^j}{\sum_j s(x_t, x_j)} \in \Delta^{n-1} \qquad (1)$$

where $p^j \in \Delta^{n-1}$ is the unit vector assigning probability 1 to ω_j.

Observe that (unqualified) empirical frequencies (of states in Ω) constitute a special case of this formula, where the function s is constant. Another special case is given by $s(x_t, x_j) = 1_{\{x_t = x_j\}}$.[1] In this case, (1) boils down to the empirical frequencies (of states in Ω) in the sub-database defined by x_t. Thus, formula (1) may be viewed as offering a continuous spectrum between the unconditional empirical frequencies and conditional empirical frequencies given x_t.

In this paper we study the probability assignment problem axiomatically. We consider the relationship between various databases, modeled as sequences of cases, and the probabilities they induce. We impose two axioms on the probability assignment function. The first, *invariance*, states that the order of cases in the database is immaterial. This axiom is not very restrictive if the description of a case is informative enough, including, for instance, the time of occurrence of the case. The second axiom, *concatenation*, requires that, for every two databases, the probability induced by their concatenation is a convex combination of the probabilities induced by each of them separately. In behavioral terms, this axiom states that, if each of two databases induces a preference for one act over another, then the same preference will be induced by their concatenation. Under a minor additional condition, these two axioms are equivalent to the existence of a similarity function such that the assignment of probabilities is done as a similarity-weighted average of the probabilities induced by single cases. Two additional assumptions then yield the representation (1).

In our theorem, the function s is derived from presumably observable probability assignments given various possible databases. We interpret this function as a similarity function. Yet, it need not satisfy any particular

[1] We assumed that the function s is strictly positive. This simplifies the analysis as one need not deal with vanishing denomintaors. Yet, for the purposes of the present discussion it is useful to consider the more general case, allowing zero similarity values. This case is not axiomatized in this paper.

properties, and may not even be symmetric. One may impose additional conditions, as in Billot, Gilboa, and Schmeidler (2004), under which there exists a norm $\mathbf{n}$ on $\mathbb{R}^k$ such that

$$s(x_t, x_j) = e^{-\mathbf{n}(x_t - x_j)}. \tag{2}$$

Such a function s satisfies symmetry and multiplicative transitivity (that is, $s(x, z) \geq s(x, y)s(y, z)$ for all x, y, z).[2]

The Bayesian approach calls for the assignment of a prior probability measure to a state space, and for the updating of this prior by Bayes's law given new information. Ramsey (1931), de Finetti (1937), Savage (1954), and Anscombe and Aumann (1963) provided compelling axiomatizations that justify the Bayesian approach from a normative viewpoint. But these axiomatizations do not help a predictor to form a prior if she is does not already have one. In this context, our approach can be viewed as providing a belief-generation tool that may be an aid to a predictor who wishes to develop a Bayesian prior.

Such a predictor may be convinced by our axiomatization that, in certain situations, it might be desirable to generate beliefs according to formula (1). Yet, just as Bayesian axiomatizations do not serve to choose a prior, our axiomatization does not provide help in choosing the similarity function. Even if one adopts a certain functional form as in (2), the question still remains, which specific similarity function should we choose?

We believe that this question is, in the final analysis, an empirical one. Hence, the similarity function should be estimated from past data. Gilboa, Lieberman, and Schmeidler (2004) axiomatize formula (1) for the case $n = 2$ (not dealt with in this paper), and develop the statistical theory required for the estimation of the function s, assuming that such a function governs

[2]Billot, Gilboa, and Schmeidler (2004) deal with a similarity-weighted average for a single real-valued variable, assuming that values of the same variables were observed in the past. Their axioms may be applied to any single component of the probability vector discussed here.

the data generating process. The present paper provides an axiomatization for the case $n > 2$. In certain situations, it allows to reduce the question of belief formation to the problem of similarity assessment. Developing the corresponding statistical theory is beyond the scope of this paper.

2 Model and Result

Let $\Omega = \{1, ..., n\}$ be a set of *states of nature*, $n \geq 3$.[3] Let C be a non-empty set of *cases*. C may be an abstract set of arbitrarily large cardinality. A *database* is a sequence of cases, $D \in C^r$ for $r \geq 1$. The set of all databases is denoted $C^* = \cup_{r \geq 1} C^r$. The concatenation of two databases, $D = (c_1, ..., c_r) \in C^r$ and $E = (c'_1, ..., c'_t) \in C^t$ is denoted by $D \circ E$ and it is defined by $D \circ E = (c_1, ..., c_r, c'_1, ..., c'_t) \in C^{r+t}$.

Observe that the same element of C may appear more than once in a given database. This structure implicitly assumes that additional observations of the same case do in fact add information. Indeed, when one estimates probabilities by relative frequencies, one subscribes to the same assumption.

For the statement of our main result we need not assume that C and Ω are a-priori related. We therefore impose no structure on C, simplifying notation and obtaining a more general result. Yet, the intended interpretation is as in the Introduction, namely, that C is a subset of $\mathbb{R}^k \times \Omega$. The prediction problem at hand, described above by $x_t \in \mathbb{R}^k$, is fixed throughout this discussion. We therefore suppress it from the notation when no confusion is likely to arise.

For each $D \in C^*$, the predictor has a probabilistic belief $p(D) \in \Delta(\Omega)$ about the realization of $\omega \in \Omega$ in the problem under discussion.

For $r \geq 1$, let Π_r be the set of all permutations on $\{1, ..., r\}$, i.e., all

[3]Our result only holds when the range of the probability assignment function is not contained in a line segment. The condition $n \geq 3$ is obviously a necessary but insufficient condition for this requirement to hold. We mention it here in order to highlight the fact that the case $n = 2$ is not covered by our result. See Gilboa, Lieberman, and Schmeidler (2004).

bijections $\pi : \{1, ..., r\} \rightarrow \{1, ..., r\}$. For $D \in C^r$ and a permutation $\pi \in \Pi_r$, let πD be the permuted database, that is, $\pi D \in C^r$ is defined by $(\pi D)_i = D_{\pi(i)}$ for $i \leq r$.

We formulate the following axioms.

Invariance: For every $r \geq 1$, every $D \in C^r$, and every permutation $\pi \in \Pi_r$, $p(D) = p(\pi D)$.

Concatenation: For every $D, E \in C^*$, $p(D \circ E) = \lambda p(D) + (1-\lambda) p(E)$ for some $\lambda \in (0, 1)$.

The Invariance axiom might appear rather restrictive, as it does not allow cases that appear later in D to have a greater impact on probability assessments than do cases that appear earlier. But this does not mean that cases that are chronologically more recent cannot have a greater weight than less recent ones. Indeed, should one include time as one of the variables in X, all permutations of a sequence of cases would contain the same information. In general, cases that are not judged to be exchangeable differ in values of some variables. Once these variables are brought forth, the Invariance axiom seems quite plausible.

The Concatenation axiom states that the beliefs induced by the concatenation of two databases cannot lie outside the interval connecting the beliefs induced by each database separately. If an expected payoff maximizer is faced with a decision problem where the states of nature are Ω, the Concatenation axiom could be re-stated as follows: for every two acts a and b, if a is (weakly) preferred to b given database D as well as given database E, then a is (weakly) preferred to b given the database $D \circ E$, and a strict preference given one of $\{D, E\}$ suffices for a strict preference given $D \circ E$.

We can now state our main result.

Theorem 1 *Let there be given a function $p : C^* \rightarrow \Delta(\Omega)$. The following are equivalent:*

(i) p satisfies the Invariance axiom, the Combination axiom, and not all

$\{p(D)\}_{D\in C^*}$ are collinear;

(ii) There exists a function $\hat{p} : C \to \Delta(\Omega)$, where not all $\{\hat{p}(c)\}_{c\in C}$ are collinear, and a function $s : C \to \mathbb{R}_{++}$ such that, for every $r \geq 1$ and every $D = (c_1, ..., c_r) \in C^r$,

$$p(D) = \frac{\sum_{j\leq r} s(c_j)\hat{p}(c_j)}{\sum_{j\leq r} s(c_j)}. \qquad (*)$$

Moreover, in this case the function $\hat{p}$ is unique, and the function s is unique up to multiplication by a positive number.

This theorem may be extended to a general measurable state space Ω with no additional complications, because for every D only a finite number of measures are involved in the formula for $p(D)$.

Theorem 1 deals with an abstract set of cases C. Let us now assume, as in the Introduction, that a case c_j is a $(k+1)$-tuple $(x_j, \omega_j) \in \mathbb{R}^k \times \Omega$, and that the function p is defined for every database D, and a given point $x_t \in \mathbb{R}^k$. The theorem then states that, under the non-collinearity condition, a function $p(D) = p(x_t, D)$ on C^* satisfies the Invariance and Concatenation axioms if and only if there are functions $s(c_j) = s(x_t, c_j)$ and $\hat{p}(c_j) = \hat{p}(x_t, c_j)$ on C such that $(*)$ holds for $p(D) = p(x_t, D)$.

This application of formula $(*)$ is more general than formula (1) in two ways: first, $\hat{p}(x_t, c_j)$ need not equal p^j, namely, the unit vector assigning probability 1 to state ω_j. Second, $s(x_t, c_j)$ may depend on ω_j and not only on (x_t, x_j). To obtain the representation (1), one therefore needs two additional assumptions. First, assume that a state ω that has never been observed in the database is assigned probability zero. This guarantees that $\hat{p}(x_t, c_j) = p^j$. Second, assume that if the names of the states of nature are permuted in the entire database, then the resulting probability vector is accordingly permuted. This would guarantee the independence of $s(x_t, c_j)$ of ω_j.

Limitations

Formula (1) might be unreasonable when the entire database is very small. Specifically, if there is only one observation, resulting in state ω_i, p_t assigns probability 1 to ω_i for any x_t. This appears to be quite extreme. However, for large databases it may be acceptable to assign zero probability to a state that has never been observed. Moreover, a state that has never been observed may not be conceived of to begin with. That is, for many applications it seems natural to define Ω as the set of states that have been observed in the past. In this case, (1) assigns a positive probability to each state.

The intended application of formula (1) is for the assignment of probabilities given databases that are large, but that are not large enough to condition on every possible combination of values of $(X^1, ..., X^k)$. Indeed, one may assume that the function p is defined only on a restricted domain of large databases, such as $C_L^* = \cup_{n \geq L} C^n$ for a large $L \geq 1$. It is straightforward to extend our result to such restricted domains.

The Concatenation axiom that we use in this paper is very similar in spirit to the Combination axiom used in Gilboa and Schmeidler (2003). Much of the discussion of this axiom in that paper applies here as well. In particular, there are two important classes of examples wherein the Concatenation axiom does not seem plausible. The first includes situations where the similarity function is learnt from the data.[4] The second class of examples involves both inductive and deductive reasoning. For instance, if we try to learn the parameter of a coin, and then use this estimate to make predictions over several future tosses, the Concatenation axiom is likely to fail.

3 Appendix: Proof

It is obvious that (ii) implies the Invariance axiom. Hence we may restrict attention to functions p that satisfy the Invariance axiom, and show that for such functions, (ii) is equivalent to the Concatenation axiom combined with

[4]The estimation procedure in Gilboa, Liebermen, and Schmeidler (2004) estimates the similarity function from the data, but assumes that these data were generated according to a *fixed* (though unknown) similarity function. However, when the data generating process itself involves an evolving similarity function, our formulae and estimation procedures are no longer valid.

the condition that not all $\{p(D)\}_{D\in C^*}$ are collinear.

In light of the Invariance axiom, a database $D \in C^*$ can be identified with a counter vector $I_D : C \to \mathbb{Z}_+$, where $I_D(c)$ is the number of times that c appears in D. Formally, for $D = (c_1, ..., c_r)$ let $I_D(c) = \#\{\, i \le r \,|\, c_i = c \,\}$. The set of counter vectors obtained from all databases $D \in C^*$ is $\mathcal{I} = \{\, I : C \to \mathbb{Z}_+ \,|\, 0 < \sum_{j\in C} I(j) < \infty \,\}$. For $I \in \mathcal{I}$, define $p(I) = p(D)$ for a $D \in C^*$ such that $I = I_D$. It is straightforward that for each $I \in \mathcal{I}$ such a D exists, and that, due to the Invariance axiom, $p(D)$ is well-defined.

We now turn to state a version of our theorem for the counter vector set-up. Observe that the concatenation of two databases D and E corresponds to the pointwise addition of their counter vectors. Formally, $I_{D\circ E} = I_D + I_E$. The Concatenation axioms is therefore re-stated as the following.

Combination: For every $I, J \in \mathcal{I}$, $p(I + J) = \lambda p(I) + (1 - \lambda)p(J)$ for some $\lambda \in (0, 1)$.

Theorem 2 *Let there be given a function $p : \mathcal{I} \to \Delta(\Omega)$. The following are equivalent:*

(i) p satisfies the Combination axiom, and not all $\{p(I)\}_{I\in\mathcal{I}}$ are collinear;

(ii) There are probability vectors $\{p^j\}_{j\in C} \subset \Delta(\Omega)$, not all collinear, and positive numbers $\{s_j\}_{j\in C}$ such that, for every I,

$$p(I) = \frac{\sum_{j\in C} s_j I(j) p^j}{\sum_{j\in C} s_j I(j)}. \qquad (*)$$

Moreover, in this case the probabilities $\{p^j\}_{j\in C}$ are unique, and the weights $\{s_j\}_{j\in C}$ are unique up to multiplication by a positive number.

Observe that Theorems 1 and 2 are equivalent. We now turn to prove Theorem 2. It is straightforward to see that (ii) implies (i). Similarly, the uniqueness part of the theorem is easily verified. We therefore only prove that (i) implies (ii).

We start with the case of a finite C, say, $C = \{1, ..., m\}$.

Remark: For every $I \in \mathcal{I}$, $k \geq 1$, $p(kI) = p(I)$.

Proof: Using the fact that $p(I + J) \in [p(I), p(J)]$ inductively.[5]$\square$

This Remark allows an extension of the domain of p to rational-coordinate vectors. Specifically, given $I \in \mathbb{Q}_+^C$, choose k such that $kI \in \mathbb{Z}_+^C$, and define $p(I)$ as identical to $p(kI)$. The Remark guarantees that the selection of k is immaterial. It follows that one may restrict attention to $p(I)$ only for $I \in \mathbb{Q}_+^C \cap \Delta(C)$, that is, for rational points in the simplex of the case types. Restricted to this domain, p is a mapping from $\mathbb{Q}_+^C \cap \Delta(C)$ into $\Delta(\Omega)$. We now state an auxiliary result that will complete the proof of (ii).[6]

Proposition 3 *Assume that* $p : \mathbb{Q}_+^m \cap \Delta^{m-1} \to \Delta^{n-1}$ *satisfies the following conditions: (i) for every* $q, q' \in \mathbb{Q}_+^m \cap \Delta^{m-1}$, *and every rational* $\alpha \in (0, 1)$, $p(\alpha q + (1 - \alpha)q') = \lambda p(q) + (1 - \lambda)p(q')$ *for some* $\lambda \in (0, 1)$*; and (ii) not all* $\{p(q)\}_{q \in \mathbb{Q}_+^m \cap \Delta^{m-1}}$ *are collinear. Then there are probability vectors* $\{p^j\}_{j \leq m} \subset \Delta^{n-1}$*, not all of which are collinear, and positive numbers* $\{s_j\}_{j \leq m}$ *such that, for every* $q \in \mathbb{Q}_+^m \cap \Delta^{m-1}$,

$$p(q) = \frac{\sum_{j \leq m} s_j q_j p^j}{\sum_{j \leq m} s_j q_j} \qquad (\bullet).$$

Proof.

For $j \leq m$, let q^j denote the j-unit vector in $\mathbb{R}^m$, i.e., the j-th extreme point of Δ^{m-1}. Obviously, one has to define $p^j = p(q^j)$. Observe that, since $p(\alpha q + (1 - \alpha)q')$ is a convex combination of $p(q)$ and $p(q')$, not all $\{p(q^j) = p^j\}_{j \leq m}$ are collinear.

We have to show that there are positive numbers $\{s_j\}_{j \leq m}$ such that $(\bullet)$ holds for every $q \in \mathbb{Q}_+^m \cap \Delta^{m-1}$.

Step 1: $m = 3$.

[5] Throughout this paper, the interval defined by two vectors, p and q, is given by $[p, q] = \{\, \lambda p + (1 - \lambda)q \mid \lambda \in [0, 1] \,\}$.

[6] The following proposition is a manifestation of a general principle, stating that functions that map intervals onto intervals are projective mappings. Another manifestation of this principle in decision theory can be found in Chew (1983).

Let $q^* = \frac{1}{3}(q^1 + q^2 + q^3)$. Choose positive numbers s_1, s_2, s_3 such that ($\bullet$) holds for q^*. Observe that such s_1, s_2, s_3 exist and are unique up to multiplication by a positive number. Define $p_s(q) = \frac{\sum_{j \leq m} s_j q_j p^j}{\sum_{j \leq m} s_j q_j}$ for all $q \in \mathbb{Q}^3_+ \cap \Delta^2$. Denote $E = \{\, q \in \mathbb{Q}^3_+ \cap \Delta^2 \,|\, p_s(q) = p(q) \,\}$. We know that $\{q^1, q^2, q^3, q^*\} \subset E$, and we wish to show that $E = \mathbb{Q}^3_+ \cap \Delta^2$.

Step 1.1: Simplicial points are in E:

The first simplicial partition of $\mathbb{Q}^3_+ \cap \Delta^2$ is a partition to four triangles separated by the segments connecting $\{(\frac{1}{2}q^1 + \frac{1}{2}q^2), (\frac{1}{2}q^2 + \frac{1}{2}q^3), (\frac{1}{2}q^3 + \frac{1}{2}q^1)\}$. The second simplicial partition is obtained by similarly partitioning each of the four triangles to four smaller triangles, and the k-th simplicial partition is defined recursively. The simplicial points of the k-th simplicial partition are all the vertices of triangles of this partition.

We now state the following

Claim: If the vertices and the center of gravity of a simplicial triangle are in E, then so are the vertices and center of gravity of all of its four simplicial sub-triangles.

Proof:

Insert Figure 1 Here

If four points that are not collinear, a, b, c, d, are in E, then the point defined by the intersection of the segments $[a, b]$ and $[c, d]$ is also in E. The proof is conducted by applying this fact inductively as suggested by Figure 1.

Explicitly, let $\{q_k^1, q_k^2, q_k^3\}$ be the vertices of a triangle in the k-th simplicial partition. Assume that $q_k^1, q_k^2, q_k^3, \frac{1}{3}(q_k^1 + q_k^2 + q_k^3) \in E$. We first show that $(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2), (\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3), (\frac{1}{2}q_k^3 + \frac{1}{2}q_k^1) \in E$. Indeed, $(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2)$ is the intersection of the line connecting q_k^3 and $\frac{1}{3}(q_k^1 + q_k^2 + q_k^3)$, and the line connecting q_k^1 and q_k^2. Hence both $p(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2)$ and $p_s(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2)$ have to be the intersection of the line connecting $p(q_k^3) = p_s(q_k^3)$ and $p(\frac{1}{3}(q_k^1 + q_k^2 + q_k^3)) = p_s(\frac{1}{3}(q_k^1 + q_k^2 + q_k^3))$, and the line connecting $p(q_k^1) = p_s(q_k^1)$ and $p(q_k^2) = p_s(q_k^2)$. Since not all $p(q)$ are collinear, this intersection is unique. Hence $(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2) \in E$. Similarly, we also have $(\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3), (\frac{1}{2}q_k^3 + \frac{1}{2}q_k^1) \in E$.

Next consider the center of gravity of the four sub-triangles. For the triangle $conv\{(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2), (\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3), (\frac{1}{2}q_k^3 + \frac{1}{2}q_k^1)\}$, the center of gravity is equal to that of $conv\{q_k^1, q_k^2, q_k^3\}$, which is already known to be in E. Next consider the center of gravity of one of the three sub-triangles that have a vertex is common with $conv\{q_k^1, q_k^2, q_k^3\}$. Assume, without loss of generality, that it is the triangle defined by $\{q_k^3, (\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3), (\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3)\}$. We first note that $\frac{1}{2}(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3) + \frac{1}{2}(\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3)$ is in E because it is the intersection of $[q^3, (\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2)]$ and $[(\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3), (\frac{1}{2}q_k^3 + \frac{1}{2}q_k^1)]$. Similarly, $\frac{1}{2}(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3) + \frac{1}{2}(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2)$ is in E. The point $\frac{1}{2}q_k^3 + \frac{1}{2}(\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3) = \frac{3}{4}q_k^3 + \frac{1}{4}q_k^2$ is on the line connecting $\frac{1}{2}(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3) + \frac{1}{2}(\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3)$ and $\frac{1}{2}(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3) + \frac{1}{2}(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^2)$ and on the line connecting q_k^2 and q_k^3. Hence $\frac{3}{4}q_k^3 + \frac{1}{4}q_k^2$ is in E. The center of gravity of the triangle $conv\{q_k^3, (\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3), (\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3)\}$ is the intersection of $[q^3, \frac{1}{2}q_k^1 + \frac{1}{2}q_k^2]$ and $[(\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3), (\frac{3}{4}q_k^3 + \frac{1}{4}q_k^2)]$. Hence the center of gravity of the triangle $conv\{q_k^3, (\frac{1}{2}q_k^1 + \frac{1}{2}q_k^3), (\frac{1}{2}q_k^2 + \frac{1}{2}q_k^3)\}$ is in E.□

Applying the claim inductively, we conclude that E contains all points that are vertices of simplicial sub-triangles of $conv\{q_k^1, q_k^2, q_k^3\}$. □

Step 1.2: Completion:

Observe that, if $q \in \mathbb{Q}_+^3 \cap conv(q, q', q'')$, then $p(q) \in conv(p(q), p(q'), p(q''))$. Consider an arbitrary $q \in conv\{q^1, q^2, q^3\}$. Take a sequence of simplicial triangles, $conv\{q_k^1, q_k^2, q_k^3\}$, such that $q \in conv\{q_k^1, q_k^2, q_k^3\}$ and that $\lim_{k\to\infty} q_k^j = q$ for all $j = 1, 2, 3$. Since p_s is a continuous function, $\lim_{k\to\infty} p_s(q_k^j) = p_s(q)$ for all $j = 1, 2, 3$. Moreover, because both p and p_s satisfy the Combination axiom, it follows that $p(q), p_s(q) \in conv\{p(q_k^1) = p_s(q_k^1), p(q_k^2) = p_s(q_k^2), p(q_k^3) = p_s(q_k^3)\}$. This is possible only if $p(q) = p_s(q)$. Hence $q \in E$. Since the choice of q was arbitrary, $E = \mathbb{Q}_+^3 \cap \Delta^2$.

Step 2: $m > 3$.

Step 2.1: Defining s_j:

Consider a triple $j, k, l \leq m$ such that $\{p^j, p^k, p^l\}$ are not collinear. Apply Step 1 to obtain a representation

$$p(q) = \sum_{\nu \in \{j,k,l\}} s_\nu^{\{j,k,l\}} q_\nu p^\nu(\{j,k,l\}) / \sum_{\nu \in \{j,k,l\}} s_\nu^{\{j,k,l\}} q_\nu$$

for all $q \in \mathbb{Q}_+^m \cap conv(\{q^j, q^k, q^l\})$. Moreover, for all $\nu \in \{j,k,l\}$, $p^\nu(\{j,k,l\}) = p(q^\nu) = p^\nu$, and the coefficients $\{s_\nu^{\{j,k,l\}}\}_{\nu \in \{j,k,l\}}$ are unique up to multiplication by a positive number.

Next consider all triples $j, k, l \leq m$ such that $\{p^j, p^k, p^l\}$ are not collinear. We argue that, for given j, k, $s_j^{\{j,k,l\}} / s_k^{\{j,k,l\}}$ is independent of l. To see this, assume that l and l' are such that neither $\{p^j, p^k, p^l\}$ nor $\{p^j, p^k, p^{l'}\}$ are collinear. Restricting attention to rational combinations of q^j and q^k, one observes that $s_j^{\{j,k,l\}} / s_k^{\{j,k,l\}} = s_j^{\{j,k,l'\}} / s_k^{\{j,k,l'\}}$. Denote this ratio by γ_{jk}. Observe that it is defined for every distinct $j, k \leq m$, because for every j, k there exists at least one l such that $\{p^j, p^k, p^l\}$ are not collinear. Further, note that if $\{p^j, p^k, p^l\}$ are not collinear, then $\gamma_{jk}\gamma_{kl}\gamma_{lj} = 1$.

Define $s_1 = 1$ and $s_j = \gamma_{j1}$ for $1 < j \leq m$. We wish to show that, for every triple $j, k, l \leq m$ such that $\{p^j, p^k, p^l\}$ are not collinear, $\{s_\nu^{\{j,k,l\}}\}_{\nu \in \{j,k,l\}}$ is proportional to $\{s_j, s_k, s_l\}$. Without loss of generality, it suffices to show that $s_j^{\{j,k,l\}} / s_k^{\{j,k,l\}} = s_j / s_k$, or that $\gamma_{jk} = s_j / s_k$. If $\{p^1, p^j, p^k\}$ are not collinear, then this equation follows from $\gamma_{1j}\gamma_{jk}\gamma_{k1} = 1$. If, however, $\{p^1, p^j, p^k\}$ are collinear, then $\{p^1, p^j, p^l\}$ and $\{p^1, p^k, p^l\}$ are not collinear. Hence $\gamma_{kl} = s_k / s_l$ and $\gamma_{lj} = s_l / s_j$. In this case, $\gamma_{jk} = 1/\gamma_{kl}\gamma_{lj} = s_j / s_k$.

Given $s = (s_j)_{j \leq m}$, define $p_s(q) = \frac{\sum_{j \leq m} s_j q_j p^j}{\sum_{j \leq m} s_j q_j}$. Thus, we wish to show that $p(q) = p_s(q)$ for all $q \in \mathbb{Q}_+^m \cap \Delta^{m-1}$.

Step 2.2: Completion:

We prove the following claim by induction on k, $3 \leq k \leq m$:

Claim: For every subset $K \subset \{1, ..., m\}$ with $|K| = k$, if $\{p^j\}_{j \in K}$ are not collinear, then $p(q) = p_s(q)$ holds for every $q \in \Delta_K \equiv \mathbb{Q}_+^m \cap conv(\{\, q^j \mid j \in K \,\})$.

Proof: The case $k = 3$ was proven in Step 1. We assume that the claim is correct for $k \geq 3$, and we prove it for $k + 1$. Let there be given $K \subset \{1, ..., m\}$ with $|\,K\,| = k + 1$, such that $\{p^j\}_{j \in K}$ are not collinear. Let $J = \{\, j \in K \,|\, \{p^l\}_{l \in K \backslash \{j\}}$ are not collinear$\}$. Observe that, for every $j \in J$, $p(q) = p_s(q)$ holds for every $q \in \Delta_{K \backslash \{j\}}$.

We argue that $|\,J\,| \geq k$. To see this, assume that there were two distinct elements j and k, in $K \backslash J$. Then all $\{p^l\}_{l \neq j}$ are collinear, as are all $\{p^l\}_{l \neq k}$. Since $|\,K\,| = k + 1 \geq 4$, there are at least two distinct elements in $K \backslash \{j, k\}$. Both p^j and p^k are collinear with $\{p^l\}_{l \neq j,k}$, and it follows that all $\{p^l\}_{l \in K}$ are collinear, a contradiction.

Consider a rational point $q \in \mathbb{Q}^m_+$ in the relative interior of $conv(\{\, q^l \,|\, l \in K\,\})$. Denote $q = \sum_{l \in K} \alpha_l q^l$ with $\alpha_l > 0$. For every $j \in J$, Let $q(j)$ be the point in $conv(\{\, q^l \,|\, l \in K \backslash \{j\}\})$ that is on the line connecting q^j and q, that is, $q(j) = \sum_{l \in K \backslash \{j\}} \frac{\alpha_l}{1 - \alpha_j} q^l$. Obviously, $p_s(q^j) = p(q^j) = p^j$. Moreover, since $j \in J$, one may apply the claim to $K \backslash \{j\}$, yielding $p_s(q(j)) = p(q(j))$. Since p_s satisfies the Combination axiom, it follows that both $p(q)$ and $p_s(q)$ are on the interval $[p_s(q^j), p_s(q(j))] = [p^j, p(q(j))]$.

Next we wish to show that, for at least two elements $j, k \in J$, the intervals $[p^j, p(q(j))]$ and $[p^k, p(q(k))]$ cannot lie on the same line. Assume not, that is, that all intervals $\{[p^j, p(q(j))]\}_{j \in J}$ lie on a line L. If $J = K$, this implies that all $\{p^j\}_{j \in K}$ are collinear, a contradiction. Assume, then, that there is an i such that $J = K \backslash \{i\}$. In this case, p^i is not on L. For $j \in J$, consider $q(j)$ as a convex combination of q^i and a point $q' \in conv(\{\, q^l \,|\, l \in K \backslash \{i, j\}\,\})$. By the Combination axiom, $p(q')$ is on the line L. Moreover, since $p^i \neq p(q')$, $p(q(j))$ is in the open interval $(p^i, p(q'))$, and therefore not on L. But this contradicts the assumption that all intervals $\{[p^j, p(q(j))]\}_{j \in J}$ lie on L.

It follows that there are distinct $j, k \in J$ for which the intervals $[p^j, p(q(j))]$ and $[p^k, p(q(k))]$ do not lie on the same line. Hence these intervals can intersect in at most one point. Since both $p(q)$ and $p_s(q)$ are on both intervals, $p(q) = p_s(q)$ follows.

We conclude that $p(q) = p_s(q)$ holds for every rational q in the relative interior of $conv(\{ q^j \mid j \in K \})$, as well as for all rational points in $conv(\{ q^l \mid l \in K \setminus \{j\}\})$ for $j \in J$. It is left to show that $p(q) = p_s(q)$ for rational points in $conv(\{ q^l \mid l \in K \setminus \{i\}\})$ for $i \in K \setminus J$. Assume not. Then, for some $q \in \mathbb{Q}_+^m \cap conv(\{ q^l \mid l \in K \setminus \{i\}\})$, $p(q) \neq p_s(q)$. But $p(q^i) = p_s(q^i) = p^i$. Hence the interval (q^i, q) is mapped by p into $(p^i, p(q))$ and by p_s – into $(p^i, p_s(q))$. Note that these two open intervals are disjoint. But for any $q' \in (q^i, q)$ we should have $p(q') = p_s(q')$, a contradiction. $\square$

It is left to complete the proof of the sufficiency of the Combination axiom in case C is infinite. For every $B \subset C$, let $\mathcal{I}_B$ be the set of databases $I \in \mathcal{I}$ such that $\sum_{j \notin B} I(j) = 0$. For every $j \in C$, define p^j by $p(I_j)$ where I_j is defined by $I_j(j) = 1$ and $I_j(k) = 0$ for $k \neq j$. For every finite $B \subset C$, for which not all $\{p^j\}_{j \in B}$ are collinear, there is a function s_B such that $(*)$ holds for every $I \in \mathcal{I}_B$. Moreover, this function is unique up to multiplication by a positive number. Fix one such finite set C_0 and choose a function s_{C_0}. For every other finite $B \subset C$, for which not all $\{p^j\}_{j \in B}$ are collinear, consider $B' = C_0 \cup B$. Over B' there exists a unique $s_{B'}$ that satisfies $(*)$ for all $I \in \mathcal{I}_{B'}$ and that extends s_{C_0}. Define s_B as the restriction of $s_{B'}$ to B. To see that this construction is well-defined, suppose that B_1 and B_2 are two such sets with a non-empty intersection. Consider $B = B_1 \cup B_2$. Since s_{B_1} and s_{B_2} are both restrictions of s_B, they are equal on $B_1 \cap B_2$. $\square\square$

References

[1] Anscombe, F. J. and R. J. Aumann (1963), "A Definition of Subjective Probability", *The Annals of Mathematics and Statistics*, **34**: 199-205.

Billot, A., I. Gilboa, and D. Schmeidler (2004), "An Axiomatization of an Exponential Similarity Function", mimeo.

Chew, S. H. (1983), "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox", *Econometrica*, **51**: 1065-1092.

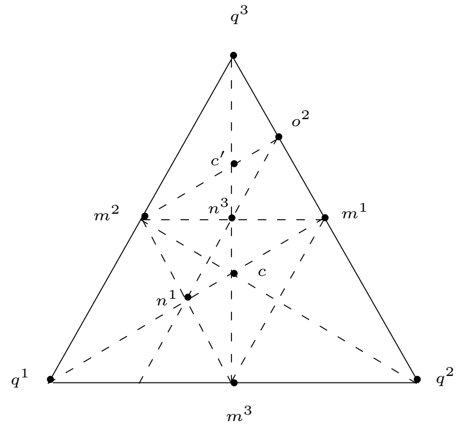
de Finetti, B. (1937), "La Prevision: Ses Lois Logiques, Ses Sources Subjectives", *Annales de l'Institute Henri Poincare*, **7**: 1-68.

Gilboa, I. and D. Schmeidler (2003), "Inductive Inference: An Axiomatic Approach", *Econometrica*, **71**: 1-26.

Gilboa, I., O. Lieberman, and D. Schmeidler (2004), "Empirical Similarity", mimeo.

Ramsey, F. P. (1931), "Truth and Probability", *The Foundation of Mathematics and Other Logical Essays*. New York: Harcourt, Brace and Co.

Savage, L. J. (1954), *The Foundations of Statistics*. New York: John Wiley and Sons.



The point m^1 is the intersection of the lines q^2q^3 and q^1c. The points m^2 and m^3 are similarly constructed. The point n^3 is the intersection on m^1m^2 and q^3c. The point n^1 is similarly constructed. The point o^2 is the intersection of n^1n^3 and q^2q^3. Finally, the center of gravity of $m^1m^2q^3$ is the intersection of m^2o^2 and q^3m^3 at c'.

Figure 1: The vertices and center of gravity of four sub-triangles

4 Appendix for Referees: Compatibility with Bayesianism

Whether our approach is compatible with Bayesianism depends on the exact definition of the state space. In our problem, there are at least three levels at which the Bayesian approach may be applied. The first, which is minimal in terms of its information requirements, is to suggest that the physician have a prior probability measure on the product space, $\Theta \equiv \mathbb{R}^k \times \Omega$, describing the joint distribution of all the variables $(X^1, ..., X^k)$ as well as the possible outcomes, Ω. If such a prior exists, all that the physician needs to do is to update this prior, given the values of $(X^1, ..., X^k)$, and obtain a posterior on Ω.

Our similarity-weighted relative frequencies can be viewed as a step towards the generation of a prior over Θ. Specifically, formula (1) suggests a method for the generation of beliefs over Ω given every possible combination of values for $(X^1, ..., X^k)$. If these posteriors are coupled with some marginal over $(X^1, ..., X^k)$, a prior over Θ will result. For prediction of the state $\omega \in \Omega$ given $(X^1, ..., X^k)$, a complete prior is not necessary. Yet, our approach is consistent with the Bayesian approach, as applied to Θ, and may be viewed as complementary to it.

At the other extreme, one may apply the Bayesian approach to a much more informative state space, allowing all conceivable observations without imposing any additional structure on the problem. This would mean that the physician has a prior distribution over all the *sequences* of observations she may obtain. Thus, the state space is $\Psi = \cup_{t \geq 1} \Theta^t$, and for every t the prior induces a well-defined marginal distribution on Θ^t. This marginal distribution can be updated given $(t-1)$ past observations, as well as the t-th realization of $(X^1, ..., X^k)$, and the posterior on ω_t can be computed.

This application of the Bayesian approach would result in beliefs over Ω generated by Bayes's updating. It is not clear, however, how one should generate a prior belief over the much larger state space $\Psi = \cup_{t\geq 1}\Theta^t$. Past observations can hardly provide the required information, since such observations are already included in the Ψ, whereas the prior should reflect the beliefs one has *before* obtaining these observations. At any rate, our approach in consistent with the Bayesian approach when applied to the space Ψ. Indeed, the only constraint imposed by the Bayesian approach at this level is the following "sure-thing principle": the posterior on Ω given a sequence of observations of length t is a weighted average of the posteriors given all possible continuations of the sequence of length $(t+1)$. (See Green and Park (1996).) It is readily observed that our formula satisfies this constraint.

The standard application of the Bayesian approach in statistics is at an intermediate level: it assumes that the observations are drawn from Θ in an i.i.d. manner, but that the probability law of this process is not known. Rather, there exists a prior over a certain set of possible probability laws. This prior induces a probability over all Ψ, but it imposes additional structure on the problem.

The prior of the probability law governing the data generating process should be derived from some theory, or past instances of similar statistical problems. However, if one uses Bayesian update at the level of the probability laws, and then deduces beliefs over Ω from it, one will typically not satisfy the Concatenation axiom.[7] Hence our formula is inconsistent with this application of the Bayesian approach. Indeed, this inconsistency is apparent even if all observations in the database share their x values. In this case, our formula reduces to estimating probabilities over Ω by relative frequencies, and this method of estimation is, in general, inconsistent with Bayesian inference about the underlying probability law.[8]

Our formula proposes a method for assigning probabilities, which is an

[7] See the Limitation sub-section below.

[8] To see this, one may consider replication of the database. Such a replication does not change the relative frequencies, but it induces a higher posterior on the probability laws that maximize the likelihood function.

extension of simple relative frequencies. It is designed to deal with databases that are not homogenous, that is, that differ in their x values. But it does not attempt to deal with situations in which one can conceive of all the possible probability laws, and feel confident enough to have a prior over them.

References

[1] Green, E. and I-U. Park (1996), "Bayes Contingent Plans", *Journal of Economic Behavior & Organization*, **31**: 225-236.