Econometrica Online Article Comment

PROBABILITIES AS SIMILARITY-WEIGHTED FREQUENCIES: A COMMENT

FAN WANG
HEC Paris

This note provides supplementary arguments for Billot, Gilboa, Samet, and Schmeidler (2005, hereafter BGSS). The result there is strengthened so that it also holds in situations where multiple cases are mapped to the same probability assessment, a possibility allowed but unaddressed by the original paper.

KEYWORDS: Probability, similarity, case-based reasoning, relative frequencies.

LET $\Omega \equiv \{1, 2, \dots, n\}$ denote a set of states of nature, $n \geq 3$. Let C denote a set of cases, the cardinality of which can be uncountably large. A database I is a function that maps C to $\mathbb{Z}_+$ and $\sum_{j\in C} I(j) < \infty$, where $I(j)$ denotes the number of observations of case j in the database.[1] Let $\mathcal{I}$ denote the set of all possible databases. A probabilistic belief function $p : \mathcal{I} \to \Delta(\Omega)$ maps databases to probability distributions over Ω. BGSS axiomatized a class of probabilistic belief functions that encapsulates the idea of "similarity-weighted frequencies," that is, each p in this class admits a representation such that, for every $I \in \mathcal{I}$,

$$p(I) = \frac{\sum_{j\in C} s_j I(j) p^j}{\sum_{j\in C} s_j I(j)}, \tag{1}$$

where $\{p^j\}_{j\in C} \subset \Delta(\Omega)$ are not all collinear, and $\{s_j\}_{j\in C}$ are positive weights. Specifically, BGSS showed that this class is logically equivalent to $\{p(I)\}_{I\in\mathcal{I}}$ being not collinear and the following axiom (see BGSS, Theorem 1 and Theorem 2).

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

It is clear from equation (1) that p^j is the probabilistic belief when the database consists of observations of case j only. And, it is allowed for two distinct cases j and k to be associated with the same probabilistic belief, that is, $p^j = p^k$. Indeed, this possibility necessarily materializes in BGSS's motivating example (see BGSS, equation (1)), where $\{p^j\}_{j\in C}$ is a finite set yet C can be uncountably large.

However, BGSS axiomatized equation (1) with the implicit premise that distinct cases lead to distinct probability assignments. In this note, we show that the premise can be dispensed with, therefore establishing the validity of BGSS's result in its full potential. Specifically, we give an alternative proof of BGSS's Proposition 3, which implies the main result when the number of cases is finite. The proposition is quoted and proved below.

Fan Wang: fan.wang1@hec.edu
I am grateful to Itzhak Gilboa for discussion and encouragement. I thank the two reviewers, whose advice greatly improved the exposition.
[1]To streamline exposition, databases are directly modeled as counter vectors, a simplification allowed by BGSS's Invariance axiom.

 https://doi.org/10.3982/ECTA16944

PROPOSITION: *Assume that* $p : \mathbb{Q}^m_+ \cap \Delta^{m-1} \to \Delta^{n-1}$ *satisfies the 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* $\{p^j\}_{j \le m} \subset \Delta^{n-1}$, *not all of which are collinear, and positive numbers* $\{s_j\}_{j \le m}$ *such that, for every* $q \in \mathbb{Q}^m_+ \cap \Delta^{m-1}$,

$$p(q) = \frac{\sum_{j \le m} s_j q_j p^j}{\sum_{j \le m} s_j q_j}. \tag{2}$$

PROOF: Step 1: Suppose $m = 3$. This is proved in BGSS's original step 1.

Step 2: Suppose $m > 3$. For every $j \in \{1, 2, \ldots, m\}$, let q^j denote the j unit vector in $\mathbb{Q}^m_+$, that is, the jth vertex of Δ^{m-1}, and define $p^j = p(q^j)$.

We start by constructing $\{s_j\}_{j \le m}$. Without loss of generality, assume $p(q^1)$, $p(q^2)$, and $p(q^3)$ are not collinear. By Step 1, we fix a triplet $\{s_j\}_{j \le 3}$ such that equation (2) is satisfied for any $q \in \mathbb{Q}^m_+ \cap \Delta(q^1, q^2, q^3)$. For every $i \in \{4, \ldots, m\}$, we fix a $l(i) \in \{1, 2, 3\}$ such that $\{p(q^i)\} \cup \{p(q^j)\}_{j \le 3, j \ne l(i)}$ are not collinear. By Step 1 again, we choose the unique $\{s'_i\} \cup \{s'_j\}_{j \le 3, j \ne l(i)}$ such that equation (2) is satisfied on $\mathbb{Q}^m_+ \cap \Delta(\{q^i\} \cup \{q^j\}_{j \le 3, j \ne l(i)})$ and $\{s'_j\}_{j \le 3, j \ne l(i)} = \{s_j\}_{j \le 3, j \ne l(i)}$. Define $s_i = s'_i$.

With $\{s_j\}_{j \le m}$, we define $p_s(q) = \sum_{j \le m} s_j q_j p^j / \sum_{j \le m} s_j q_j$ for all $q \in \mathbb{Q}^m_+ \cap \Delta^{m-1}$, and define $E = \{q \in \mathbb{Q}^m_+ \cap \Delta^{m-1} : p(q) = p_s(q)\}$. By construction, we have $\mathbb{Q}^m_+ \cap \Delta(q^1, q^2, q^3) \subset E$ and $\mathbb{Q}^m_+ \cap \Delta(\{q^i\} \cup \{q^j\}_{j \le 3, j \ne l(i)}) \subset E$ for every $3 < i \le m$. We want to show $E = \mathbb{Q}^m_+ \cap \Delta^{m-1}$.

The proof relies on a key observation made by BGSS. Given four points $a, b, c, d \in E$ such that $p_s(a)$, $p_s(b)$, $p_s(c)$, and $p_s(d)$ are distinct, if the line connecting a to b intersects the line connecting c to d uniquely at point $e \in \mathbb{Q}^m_+ \cap \Delta^{m-1}$, and the line connecting $p_s(a)$ to $p_s(b)$ intersects the line connecting $p_s(c)$ to $p_s(d)$ uniquely, then we must have $e \in E$. Here is why. First notice that since condition (i) implies that $p(e)$ has to be on the line connecting $p_s(a)$ to $p_s(b)$ and also on the line connecting $p_s(c)$ to $p_s(d)$, the two lines' unique intersection has to be $p(e)$. Then notice that p_s maps any line in $\mathbb{Q}^m_+ \cap \Delta^{m-1}$ to either a line or a point in Δ^{n-1}, and therefore the line connecting $p_s(a)$ to $p_s(b)$ (resp. $p_s(c)$ to $p_s(d)$) is the image of the line connecting a to b (resp. c to d) under p_s. So, it must be that $p(e) = p_s(e)$.[2]

For any $i > 3$, suppose $\mathbb{Q}^m_+ \cap \Delta(q^1, q^2, \ldots, q^{i-1}) \subset E$, and we show $\mathbb{Q}^m_+ \cap \Delta(q^1, q^2, \ldots, q^i) \subset E$.

Consider any $q = \sum_{j \le i} \beta_j q^j \in \mathbb{Q}^m_{++} \cap \Delta(q^1, q^2, \ldots, q^i)$. Define

$$q(-i) = \frac{1}{\sum_{j<i} \beta_j} \sum_{j<i} \beta_j q^j \in E.$$

[2]Notice there might be other $\{s'\}_{j \le m}$ and $p_{s'}$ such that $p_s(x) = p_{s'}(x)$ for any $x \in \{a, b, c, d, e\}$.

If $p(q(-i)) = p(q^i)$, q is necessarily in E by condition (i). If $p(q(-i)) \neq p(q^i)$, for any vector of rational numbers $\alpha = \{\alpha_j\}_{j\leq 3, j\neq l(i)} \in [0,1]^2$, define

$$\hat{q}(\alpha) = \hat{r}(\alpha)\left(\beta_i q^i + \sum_{j\leq 3, j\neq l(i)} \alpha_j \beta_j q^j\right), \quad \text{and,}$$

$$\tilde{q}(\alpha) = \tilde{r}(\alpha)\left(\beta_{l(i)} q^{l(i)} + \sum_{j\leq 3, j\neq l(i)} (1-\alpha_j)\beta_j q^j + \sum_{3<j<i} \beta_j q^j\right),$$

where $\hat{r}(\alpha), \tilde{r}(\alpha) \in (1, +\infty)$ are the unique normalizers which make $\hat{q}(\alpha)$ and $\tilde{q}(\alpha)$ in $\mathbb{Q}^m_+ \cap \Delta(q^1, q^2, \dots, q^i)$, respectively. We have $\hat{q}(\alpha) \in E$ because $\hat{q}(\alpha) \in \mathbb{Q}^m_+ \cap \Delta(\{q^j\}_{j\leq 3, j\neq l(i)} \cup \{q^i\}) \subset E$, and, $\tilde{q}(\alpha) \in E$ because $\tilde{q}(\alpha) \in \mathbb{Q}^m_+ \cap \Delta(q^1, q^2, \dots, q^{i-1}) \subset E$ by assumption. Notice that $\frac{1}{\hat{r}(\alpha)} + \frac{1}{\tilde{r}(\alpha)} = 1$ and, in particular, $q = \frac{1}{\hat{r}(\alpha)}\hat{q}(\alpha) + \frac{1}{\tilde{r}(\alpha)}\tilde{q}(\alpha)$.

We claim that $\{\alpha_j\}_{j\leq 3, j\neq l(i)}$ can be chosen to make $p(q(-i))$, $p(q^i)$, and $p(\hat{q}(\alpha))$ not collinear, which by condition (i) also implies $q(-i)$, q^i, and $\hat{q}(a)$ being not collinear. If $p(q(-i))$, $p(q^i)$, and $p(\hat{q}(\alpha))$ are collinear for all $\{\alpha_j\}_{j\leq 3, j\neq l(i)}$, then $\{p(q^j)\}_{j\leq 3, j\neq l(i)}$, $p(q(-i))$, and $p(q^i)$ are collinear, which contradicts $\{p(q^i)\} \cup \{p(q^j)\}_{j\leq 3, j\neq l(i)}$ being not collinear. So, the claim is valid. Fix any such α. Then, q (resp. $p(q)$) is the unique intersection of the line connecting $q(-i)$ to q^i (resp. $p_s(q(-i))$ to $p_s(q^i)$) and the line connecting $\hat{q}(\alpha)$ to $\tilde{q}(\alpha)$ (resp. $p_s(\hat{q}(\alpha))$ to $p_s(\tilde{q}(\alpha))$), and therefore $q \in E$. Because q is arbitrary, we have $\mathbb{Q}^m_{++} \cap \Delta(q^1, q^2, \dots, q^i) \subset E$.

Consider $q = \sum_{j\leq i} \beta_j q^j \in \mathbb{Q}^m_+ \cap \Delta(q^1, q^2, \dots, q^i)$ with some β_j equal to 0. Pick $q', q'' \in \mathbb{Q}^m_{++} \cap \Delta(q^1, q^2, \dots, q^i)$ such that $p(q)$, $p(q')$, and $p(q'')$ are not collinear, which can be done because $p(\mathbb{Q}^m_{++} \cap \Delta(q^1, q^2, \dots, q^i))$ is of dimension at least 2. The two line segments $\{\alpha q + (1-\alpha)q' : \alpha \in (0,1) \cap \mathbb{Q}\}$ and $\{\alpha q + (1-\alpha)q'' : \alpha \in (0,1) \cap \mathbb{Q}\}$ are in E, and the two lines extending them intersect uniquely at q. Since their images under p_s intersect uniquely at $p_s(q)$, we conclude that $q \in E$.

Taken together, the above two arguments show that $\mathbb{Q}^m_+ \cap \Delta(q^1, q^2, \dots, q^i) \subset E$. By inducting on i from 4 to m, we conclude that $E = \mathbb{Q}^m_+ \cap \Delta^{m-1}$ and the proof is completed. *Q.E.D.*

REFERENCES

BILLOT, A., I. GILBOA, D. SAMET, AND D. SCHMEIDLER (2005): "Probabilities as Similarity-Weighted Frequencies," *Econometrica*, 73, 1125–1136. [1]

Co-editor Joel Sobel handled this manuscript.

Manuscript received 31 December, 2018; final version accepted 10 June, 2019.