Lexicographic Probabilities as Similarity-Weighted Frequencies

Jacob D. Leshno*

February 6, 2014

Abstract

How does a decision maker form beliefs from a database of past observations? This paper gives an axiomatization of a decision maker who filters the databases for relevant observations, and forms his belief by similarity-weighted average of the beliefs induced by each relevant observation. This result extends the axiomatization of Billot, Gilboa, Samet and Schmeidler (2005) by endogenizing the selection of relevant observations. This belief formation process naturally generates the lexicographic beliefs systems of Blume, Brandenburger, and Dekel (1991), where higher order beliefs are generated by observations that are relatively irrelevant.

Keywords: Belief formation, case based decision making, lexicographic belief systems.

JEL: D81

*Columbia University, Graduate School of Business, jleshno@columbia.edu. This paper was initially written as a part of the requirements of the M.A. degree at Tel Aviv University, under the supervision of Itzhak Gilboa and David Schmeidler, for which I am deeply grateful. I also thank Lauren Merrill, Dov Samet, the participants of the game theory seminar at Tel Aviv University and the RUD conference, and anonymous referees for helpful comments.

1 Introduction

A physician is asked to give a prognosis for a patient by assigning probabilities to possible treatment outcomes $\Omega = \{1, .., m\}$. Given a database of past observations, (x_j, ω_j) where ω_j is the outcome and $x_j \in \mathbb{R}^k$ is the characteristics of observation j, how should he form his probabilistic belief? Billot et al. (2005) suggested the similarity-weighted frequencies rule

$$p = \frac{\sum_j s(x, x_j)\delta_j}{\sum_j s(x, x_j)} \in \Delta(\Omega)$$

where $\delta_j \in \Delta(\Omega)$ gives probability 1 to ω_j and $s(x, x_j) \in \mathbb{R}_+$ is the similarity weight between observation j and the current patient. Unqualified empirical frequencies are a special case of this rule, by setting $s(\cdot) \equiv 1$. Another special case is given by $s(\cdot) = \begin{cases} 1 & x_j = x \\ 0 & x_j \neq x \end{cases}$, taking empirical frequencies on the restricted database of observations with identical characteristics. The similarity-weighted frequencies rule allows to learn from observations with non-identical characteristics while giving higher weight to more similar and more informative observations.

Billot et al. (2005) characterized the similarity-weighted frequencies rule for a decision maker who gives all cases a strictly positive similarity-weight. This restriction implicitly assumes that the database consists only of relevant observations, but choosing which observations are relevant is in itself a question of similarity. Furthermore, the set of relevant observations may depend on the prediction problem and the database of available observations.

We endogenize the selection of relevant observations by weakening the concatenation axiom from Billot et al. (2005) — we require that the probability induced by a concatenation of two databases is a weak convex combination of the probabilities induced by each of the two, instead of requiring strict convex combination. Specifically, we allow the decision maker not to change his prediction when observations are added to his database. We give a representation theorem for the probability formation rule, which first filters irrelevant cases, and generates probabilities by a similarity-weighted average of the probabilities induced by the relevant cases.

Consider as an example a physician evaluating a patient suffering from a new cancer type. His database may contain observation on patients with same cancer type, patients with other cancer types and patients with unrelated disease. If the database includes sufficiently many observations of the same cancer type we expect the doctor to ignore all other observation. However, if observations of the same cancer type are not available we expect the doctor to learn from patients of other cancer types, while still ignoring patients of unrelated disease. Relevancy of observations depends on the database, and can be captured by a classification of observation into a hierarchy of relevancy tiers. The decision maker forms a probabilistic belief by using only observations from the most relevant tier available.

Cases from lower relevancy tiers provide additional information, which can be used when the most relevant tiers are not decisive. Consider a decision maker who is using the formed probabilities to choose between acts, which are evaluated according to an (exogenously given) expected utility function. Suppose that the probabilistic belief induced from the most relevant observations is not decisive. Rather than being indifferent, it is natural for the decision maker to use observations in lower relevancy tiers. The decision maker's belief formation process can naturally produce not only a probabilistic belief from observation in the most relevant tier, but a lexicographic probability system where a vector of probabilities is generated by a similarity-weighted average of observations in each of the available relevancy tiers. The decision maker uses higher order probabilities when indifferent without them. This yields the non-Archimedean decision maker of Blume et al. (1991a,b), which now generates his lexicographic probability system from a database of observations. By considering observation from lower relevancy tiers the decision maker can satisfy admissibility and have well defined posterior.

2 Model and Results

Let $\Omega = \{1, ..., m\}$ be a set of possible outcomes, $m \geq 3$. The decision maker is asked to give a probabilistic belief $p(D) \in \Delta(\Omega)$, based on a database D of observations. Let C be a nonempty finite set of potential observations/cases. A database is a sequence of cases, $D = (c_1, c_2, ..., c_r) \in C^r$ for $r \geq 1$.[1] The set of all databases is denoted $C^* = \cup_{r\geq 1} C^r$. Denote by $C^\circ = \{D \in C^* | \forall c \in C \ \exists i \ D_i = c\}$ the set of databases containing every case at least once. Concatenation of two databases $D = (c_1, ..., c_t) \in C^t$ and $E = (c'_1, ..., c'_r) \in C^r$, is denoted by $D \circ E = (c_1, ..., c_t, c'_1, ..., c'_r) \in C^{r+t}$. For $D \in C^r$ and a permutation $\pi : \{1, ..., r\} \to \{1, ..., r\}$, let $\pi D \in C^r$ be defined by $(\pi D)_i = D_{\pi(i)}$.
Consider the following axioms adapted from Billot et al. (2005):

INVARIANCE: For every $D \in C_r$ and a permutation π on $\{1, .., r\}$, $p(D) = p(\pi D)$.

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

Invariance states that the order of cases in the database is immaterial. Weak concatenation states that beliefs induced by the concatenation of two databases must lie in 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 weak concatenation axiom could be restated as follows: for every two acts x and y, if x is (weakly) preferred to y given database D as well as given database E, then x is (weakly) preferred to y given the database $D \circ E$. In contrast to Billot et al. (2005), a strict preference given one of $\{D, E\}$ does not necessarily imply a strict preference given $D \circ E$, as $\lambda = 1$ implies $p(D \circ E) = p(D)$.

Theorem 1. *Let $p : C^* \to \Delta(\Omega)$. The following are equivalent:*

(i) The mapping p satisfies the invariance axiom, the weak concatenation axiom, and not all $\{p(D)\}_{D \in C^\circ}$ are collinear.

[1] A case may appear more than once in a database. For example, repetition of the case "heads" documents multiple coin flips.

(ii) There exists $\hat{p} : C \to \Delta(\Omega)$, and nonnegative weights $s : C \to \mathbb{R}_+ \cup \{0\}$, where not all $\{\hat{p}(c) \; | c \in C, s(c) > 0\}$ are collinear, such that for every $D \in C^*$ for which $\sum_i s(D_i) > 0$

$$p(D) = \frac{\Sigma_i s(c_i)\hat{p}(c_i)}{\Sigma_i s(c_i)}$$

and p restricted to $\{D \in C^* | \sum_i s(D_i) = 0\}$ satisfies weak concatenation and invariance.

$s(\cdot)$ filters out irrelevant cases by giving them zero weight. The decision maker agglomerates the probabilistic predictions from relevant cases by a weighted average using weights $s(c)$. The proof, in the appendix, uses the geometric structure of $p(D)$. It constructs p by induction on the number of cases. Using that the image of p is not contained in a line, the implied geometry of p allows us to extend the representation to database with additional cases. Non-collinearity is essential, the following example shows that if the image of p is contained in a line the result fails:

Example. Let $\Omega = \{\omega_1, \omega_2\}$, $C = C_{\omega_1} \cup C_{\omega_2}$ and $p : C^* \to \Omega = \{\omega_1, \omega_2\}$. $p(D)$ is defined by $p(D) = \delta_{\omega_1}$ if a majority of cases are from C_{ω_1} and $p(D) = \delta_{\omega_2}$ otherwise. p satisfies the invariance and weak concatenation axioms, but it cannot be represented as similarity-weighted frequencies. Also notice that under p cases from C_{ω_1} are "irrelevant" in some databases but not others, showing that the classification of irrelevant cases relies on the geometric structure implied by the non-collinearity assumption.

Theorem 1 is silent for databases that include only irrelevant cases, i.e. $\sum_i s(D_i) = 0$. Let $C_2 = \{c \in C | s(c) = 0\}$ be the set of irrelevant cases, and $C_2^* = \{D \in C^* | \sum_i s(D_i) = 0\}$. If $\{p(D)\}_{D \in C_2^\circ}$ are not all collinear we can apply theorem 1 on C_2^* to derive a similarity $s_2 : C_2 \to \mathbb{R}_+ \cup \{0\}$ giving a representation for $p(D)$, $D \in C_2^*(s_2)$. Inductively, we can produce a series of similarity-weight functions $\{s_k\}$, each giving a representation of the rule for databases that the previous similarity-weights rendered irrelevant. The resulting hierarchical belief formation rule can be described by lexicographic weights. For each $c \in C$ we have a pair $(k(c), s_{k(c)}(c))$ where $k(c) = \min_k s_k(c) > 0$ is the relevancy tier. If $k(c_1) > k(c_2)$ then c_2 will be filtered in any database that includes the higher tier case c_1. Given a database, the decision maker considers only cases for which $k(c) = \min_i k(D_i)$, and aggregates them according to weights $s_{k(c)}(c)$.

A more concise "numerical" representation is given by similarity-weights that take values from $\mathbb{F}$, a non-Archimedean ordered field. $\mathbb{F}$ contains, in addition to $\mathbb{R}$, an infinitesimal $\varepsilon \in F$ that is $0 < \varepsilon < \frac{1}{n}$ for any $n \in \mathbb{N}$. The lexicographic weights $(k(c), s_{k(c)}(c))$ can be translated into a non-standard number $\varepsilon^{k(c)} s_{k(c)}(c)$. For $\sigma_1, \sigma_2 \in \mathbb{F}$ we write[2] $\sigma_1 \ll \sigma_2$ when $n\sigma_1 < \sigma_2$ for all $n \in \mathbb{N}$. To capture the selection of most relevant cases define the floor operator:

$$\lfloor \sigma_1 \rfloor_{\sigma_2} = \begin{cases} 0 & \sigma_1 \ll \sigma_2 \\ \sigma_1 & otherwise \end{cases}$$

We can now state a fuller characterization using weights from a non-Archimedean field :

Theorem 2. *Let $p : C^* \to \Delta(\Omega)$. The following are equivalent:*

(i) The function p satisfies the invariance axiom and the weak concatenation axiom.

(ii) There exists $\hat{p} : C \to \Delta(\Omega)$ and non-negative weights $s : C \to \mathbb{F}$ taken from a non-Archimedean ordered field $\mathbb{F}$, such that for every $D = (c_1, ..., c_r) \in C^*$ for which $\sum_i s(c_i) > 0$ the function p can be put as

$$p(D) = \frac{\Sigma_i \lfloor s(c_i) \rfloor_\sigma \hat{p}(c_i)}{\Sigma_i \lfloor s(c_i) \rfloor_\sigma}$$

with $\sigma = \max_{i \le r} s(c_i)$. In addition p restricted to $\tilde{C} = \{D = (c_1, ..., c_r) \in C^* | \Sigma s(c_i) = 0\}$ satisfies the weak concatenation and invariance axioms, and all $\{p(D)\}_{D \in \tilde{C}^\circ}$ are collinear.

Blume et al. (1991a) develop a non-Archimedean subjective expected utility framework which uses such a lexicographic probability system. The lexicographic proba-

[2]See Shamseddine and Berz (2010).

bility system allows the decision maker to satisfy admissibility and have well defined conditional probabilities, while allowing for "null" events. In their axiomatization they derive the lexicographic probability system from preferences, but they do not specify how one would generate such a lexicographic probability system. Our hierarchical belief formation provides a natural formation of a lexicographic probability system, where higher order beliefs are generated from irrelevant cases.

Assume our decision maker uses the probabilistic belief to choose between acts $x, y \in \Omega \to \mathcal{X}$. If the probabilities were known the decision maker would have expected utility preferences with a given utility function $u : \mathcal{X} \to \mathbb{R}$. Using $s(\cdot)$ he can naturally produce a lexicographic probability system $\vec{p}(D) = (p_1, ..., p_{K(D)}) \in \Delta^*(\Omega)$ by $p_k = p(D^k)$ where $D^1 = D$ and D^{k+1} is the database generated by erasing the highest relevancy tier cases from D^k.

Given choice between acts $x, y : \Omega \to \mathcal{X}$, and (exogenously given) utility function $u : \mathcal{X} \to \mathbb{R}$, the probabilistic belief is generated database D. If $p(D)$, using only the most relevant cases, induces indifference between acts x, y, it is natural to break indifferences using cases from lower relevancy tiers. The lexicographic preference structure follows naturally: if p_1 implies difference, $\sum_\omega p_1(D)(\omega)u(x_\omega) = \sum_\omega p_1(D)(\omega)u(y_\omega)$, remove the highest relevancy cases and form a decision from the rest. This gives the non-Archimedean subjective expected utility preferences of Blume et al. (1991a):

$$x \succcurlyeq_D y \iff \left(\sum_\omega p_k(D)(\omega)u(x_\omega) \right)_{k=1}^{K(D)} \geqslant_L \left(\sum_\omega p_k(D)(\omega)u(y_\omega) \right)_{k=1}^{K(D)}$$

where $\geqslant_L$ is the lexicographic ordering.

We thus have the following conclusion:

Remark. Suppose that $p : C^* \to \Delta(\Omega)$ satisfies the invariance and weak concatenation axioms. Then there are non-Archimedean similarity-weights $s : C \to \mathbb{F}$ that generate lexicographic probability system $\vec{p} : C^* \to \Delta^*(\Omega)$, as defined by Blume et al. (1991a). For any expected utility function u, the non-Archimedean subjective expected utility preference induced by u and $\vec{p}(D)$ is a refinement of the Archimedean subjective expected utility preference induced by u and $p(D)$.

3 Discussion and Limitations

Our weak concatenation axiom is similar to the combination axiom presented by Gilboa and Schmeidler (2003) and has the same limitations discussed there. As in kernel methods, the probability is formed under the assumption that similar cases will have similar outcome distributions. Additional assumptions on the problem can allow for induction or extrapolation. For example, a decision maker who assumes a model of a coin with a parameter probability of heads can form probabilities based on the estimated parameter, and is likely to violate our axiom. In our leading example the physician may not be able to make any further assumptions and therefore use our non-parametric rule.

In our model another difficulty may arise. Consider a large database that contains only a few cases from the highest relevancy tier. Our decision maker forms his prediction using only a small number of cases. We have no satisfying answer whether it is reasonable to limit attention to the most meaningful data when a sufficient number of observations is required for statistical analysis. However, our results do not depend on these extremes. All our results still hold when we limit the databases considered to the set $C^+ = \{D \in C^* | \forall c \in C \text{ if } \exists i \text{ s.t. } D_i = c \text{ then } \#\{i | D_i = c\} \geq N_c\}$, where a database contains a case c only if it contains a sufficient number of observations.

In order to apply the prediction formula one must have similarity-weights, but our model does not specify how should one chose the similarity function. In our model we consider cases in a general settings, assuming no structure on the cases, and thus no restrictions on the similarity function. Additional assumptions on the structure of the cases can imply specific forms of the similarity function (see Gilboa et al. (2006); Billot et al. (2008)). Our model presented can be used as means for elicitation of the similarity function in a similar way that Savage may help in the elicitation of probabilities.

References

Billot, A., I. Gilboa, and D. Schmeidler, "Axiomatization of an exponential similarity function," *Mathematical Social Sciences*, 2008, *55* (2), 107–115.

_ , _ , D. Samet, and D. Schmeidler, "Probabilities as Similarity-Weighted Frequencies," *Econometrica*, 2005, *73* (4), 1125–1136.

Blume, L., A. Brandenburger, and E. Dekel, "Lexicographic probabilities and choice under uncertainty," *Econometrica: Journal of the Econometric Society*, 1991, pp. 61–79.

_ , _ , and _ , "Lexicographic probabilities and equilibrium refinements," *Econometrica: Journal of the Econometric Society*, 1991, pp. 81–98.

Gilboa, I. and D. Schmeidler, "Inductive inference: An axiomatic approach," *Econometrica*, 2003, *71* (1), 1–26.

_ , O. Lieberman, and D. Schmeidler, "Empirical similarity," *The Review of Economics and Statistics*, 2006, *88* (3), 433–444.

Shamseddine, K. and M. Berz, "Analysis on the Levi-Civita field, a brief overview," in "Advances in p-adic and non-Archimedean analysis: tenth International Conference, June 30-July 3, 2008, Michigan State University, East Lansing, Michigan," Vol. 508 Amer Mathematical Society 2010, p. 215.

Online Appendix: Proofs

Definition and Notations

Several notations and definitions will simplify the proof. Let $|C| = n$ and $|\Omega| = m+1$ so $\Delta(C) = \Delta_{n-1}$ and $\Delta(\Omega) = \Delta_m$. $\{e_i\}_{i=1}^n$ denote the unit vectors in Δ_{n-1}, the $n-1$ dimensional simplex, thus $\Delta_{n-1} = conv(\{e_i\}_{i=1}^n)$. The face of the simplex Δ_{n-1} that is opposite the vertex e_i will be denoted $A_i = conv(\{e_1, ..., e_n\} \setminus \{e_i\})$. For a function $p : \Delta_{n-1} \to \Delta_m$, we will use the notation $p^i = p(e_i) \in \Delta_m$.

For two points $a, b \in \Delta_{n-1}$ we will denote by $[a, b] = \{\lambda a + (1-\lambda)b | 0 \leq \lambda \leq 1\}$ the segment between the points a and b. Similarly, $(a, b) = \{\lambda a + (1-\lambda)b | 0 < \lambda < 1\}$ denotes the open segment, and $(a, b] = \{\lambda a + (1-\lambda)b | 0 < \lambda \leq 1\}$ the half-open segment. If $a \neq b$ we will denote $L(x, y) = \{x + \lambda(x-y) | \lambda \in \mathbb{R}\}$ the unique line that contains the two points. The symbol $^\circ$ denotes the interior of a set, relative to the topology of the containing space. For example $[a, b]^\circ = (a, b)$ when we consider $[a, b]$ as a subset of the line $L(a, b)$, and $([a, b] \cap \mathbb{Q}^n)^\circ = (a, b) \cap \mathbb{Q}^n$ when we consider $[a, b] \cap \mathbb{Q}^n$ as a subset of the rational line $L(a, b) \cap \mathbb{Q}^n$. When it is clear, we will not explicitly state the relative containing space.

A Proof of Theorem 1

Billot et al. (2005) show that a function $p : C^* \to \Delta_m$ that satisfies the invariance axiom and the concatenation axiom depends only on the relative frequencies of the cases. Any such function can be equivalently put as a function $p : \Delta_{n-1} \cap \mathbb{Q} \to \Delta_m$ where the vectors in Δ_{n-1} represent the relative frequencies of the cases. The next definition is equivalent to the weak concatenation axiom in terms of mappings of relative frequencies.

Definition. [3] A function $p : X \to \Delta_m$, where $X \subset \Delta_{n-1}$, will be called *convex* if for

[3] Note the difference from the definition in Billot et al. (2005). They require the function to be strictly convex, i.e. $x \in (a, b) \Rightarrow p(x) \in (p(a), p(b))$. We allow for $p(x) = p(a)$ or $p(x) = p(b)$.

every $x, a, b \in X$ such that $x \in [a, b]$, p satisfies

$$p(x) \in [p(a), p(b)]$$

Our interest is in functions whose domain is $X = \Delta_{n-1} \cap \mathbb{Q}$, and as an auxiliary result we also will obtain the result for $X = \Delta_{n-1}$. For a field F denote $conv_F(\{e_1, ..., e_n\}) = \{\sum \alpha_i e_i | \sum \alpha_i = 1, \alpha_i \in F, \alpha_i \geq 0\}$. Notice that $\Delta_{n-1} \cap \mathbb{Q} = conv_{\mathbb{Q}}(\{e_1, ..., e_n\})$ and $\Delta_{n-1} = conv_{\mathbb{R}}(\{e_1, ..., e_n\})$. For any field F, $conv_F(\{e_1, ..., e_n\})$ is closed under line intersection, that is if $a, b, c, d \in conv_F(\{e_1, ..., e_n\})$ and $x \in L(a, b) \cap L(c, d)$ then $x \in conv_F(\{e_1, ..., e_n\})$. Notice also that $\Delta_{n-1} \cap \mathbb{Q}$ is dense in Δ_{n-1}. The result and all claims hold for both $\mathbb{R}$ and $\mathbb{Q}$ as well as for any $conv_F(\{e_1, ..., e_n\})$ that is dense and closed under line intersection.

For brevity of notation we will first state our claims for Δ_{n-1}. All claims and arguments remain true if Δ_{n-1} was replaced by the rational simplex $\Delta_{n-1} \cap \mathbb{Q}$. Some details need a more delicate handling for $\Delta_{n-1} \cap \mathbb{Q}$, and these will be covered in the appropriate section.

The following theorem is similar to theorem 1. It makes use of the invariance axiom and does not restrict p to rational frequencies.

Theorem 3. *Let there be given a function $p : \Delta_{n-1} \to \Delta_m$. The following are equivalent:*

(i) p is convex and $p(\Delta^\circ_{n-1})$ is not contained in any line.

(ii) There exist vectors $p^i \in \Delta_m$ and non-negative numbers $\{s_i\}_{i=1}^n$ such that $\{p^i | s_i > 0\}$ are not all collinear, and for every $x \in \Delta_{n-1}$ for which $\Sigma s_i x_i > 0$ the function p can be put as

$$p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$$

and p restricted to $\{x | \Sigma s_i x_i = 0\}$ is convex.

First we prove that (ii) implies (i). It is easy to see that (ii) implies that $p(\Delta^\circ_{n-1})$ is not contained in any line. Further, the formula for p implies that p is convex on

$\{x|\Sigma s_i x_i > 0\}$, and convexity of p on $\{x|\Sigma s_i x_i = 0\}$ is stated directly. All we need to show is that for a and b such that $\Sigma s_i a_i = 0$ and $\Sigma s_i b_i > 0$, and $c = (1-\lambda)a + \lambda b \in (a, b)$ we have $p(c) \in [p(a), p(b)]$. However, in this case we obtain that $\Sigma s_i c_i > 0$, hence[4]

$$p(c) = \frac{\Sigma s_i c_i p^i}{\Sigma s_i c_i} = \frac{(1-\lambda)\Sigma s_i a_i p^i + \lambda \Sigma s_i b_i p^i}{(1-\lambda)\Sigma s_i a_i + \lambda \Sigma s_i b_i} = \frac{\lambda \Sigma s_i b_i p^i}{\lambda \Sigma s_i b_i} = p(b)$$

and we find that $p(c) = p(b)$. In particular, $p(c) \in [p(a), p(b)]$.

We now turn to prove that $(i) \Rightarrow (ii)$. First we state some results about convex functions. We then continue to prove that $(i) \Rightarrow (ii)$ holds for the case $n = 3$, and finally continue by induction to prove this claim for $n > 3$.

Remark. One may attempt to prove Theorem 3 by first identifying the irrelevant cases, say from databases of size 2, and using the proof of Billot et al. (2005) for the relevant cases. But as example from section 2 shows, to show that irrelevant cases are irrelevant in any database in Δ_{n-1}° one must use non-collinearity and the geometric structure of p. Therefore, such a proof would have to use a construction such as the construction derived below.

Proposition 4. *$p : \Delta_{n-1} \to \Delta_m$ is convex if and only if for every convex set $C \subset \Delta_m$ $p^{-1}(C)$ is convex.*

Proof. Only if : If $a, b \in p^{-1}(C)$ and $x \in [a, b]$ then convexity of p implies $p(x) \in [p(a), p(b)]$ and convexity of C implies $[p(a), p(b)] \subset C$. Therefore $x \in p^{-1}(C)$.

If : Assume the contrary, namely, that there exists a, b and x such that $x \in [a, b]$ but $p(x) \notin [p(a), p(b)]$. Define C to be the convex set $[p(a), p(b)]$, and observe that $a, b \in p^{-1}(C)$ but $x \notin p^{-1}(C)$. Hence, $p^{-1}(C)$ isn't a convex set, a contradiction. □

Lemma 5. *If $x, y, z \in \Delta_{n-1}$ are on the same line and x and y are mapped by a convex p to different points $p(x) \neq p(y)$, then $p(z) \in L(p(x), p(y))$.*

Proof. If $z \in [x, y]$ then the lemma follows directly from the definition. Otherwise assume wlog (without loss of generality) $x \in [y, z]$. If $p(y) = p(z)$, the conclusion holds (i.e., $p(z) \in L(p(x), p(y))$). Otherwise, by convexity we find that

[4] We use the fact that if $\sum s_i a_i = 0$ then $\forall i\ s_i a_i = 0$ and thus $\sum s_i a_i p^i = 0$.

$p(x) \in L(p(z), p(y))$, we also observe that $p(x) \in L(p(x), p(y))$. Because the two lines $L(p(x), p(y))$, $L(p(z), p(y))$ intersect in two different points, $p(x), p(y)$, they must be the same line, and it follows that $p(z) \in L(p(x), p(y))$. □

Corollary 6. (constant perspective lines corollary) *Let $[x, y] \subset \Delta_{n-1}$ be a segment. If there is a line L such that $p(x) \notin L$ and $p((z, y]) \subset L$ for some $z \in [x, y]$ then p must be constant on $(z, y]$, i.e. $p((z, y]) = \{p(y)\}$. The same holds if we replace $[z, y]$ by $(z, y]$ in the above statement.*

Proof. Assume that x, y and L satisfy the conditions of the Corollary, but that p is not constant on $(z, y]$ (or $[z, y]$). Then we can find two points $z_1, z_2 \in (z, y]$ (or $z_1, z_2 \in [z, y]$) with $p(z_1) \neq p(z_2)$, and by the lemma $x \in L(p(z_1), p(z_2)) = L$ contradicting our assumption. □

Corollary 7. (intersecting lines corollary) *Given two convex mappings p, p', which are equal on $\{a_1, a_2, b_1, b_2\}$, where $L(p(a_1), p(a_2))$, $L(p(b_1), p(b_2))$ are two distinct well defined lines [5], then the mappings are also equal for $x \in [a_1, a_2] \cap [b_1, b_2]$.*

Proof. As both p and p' must map x to the single point that is both on $L(p(a_1), p(a_2))$ and $L(p(b_1), p(b_2))$ they must be equal. □

A.1 Proof for $n = 3$

We begin by proving the following lemma, which will be used in the proof. We denote $p^i = p(e_i) \in \Delta_m$.

Lemma 8. *If $p : \Delta_2 \to \Delta_2$ is convex and $p(\Delta_2^\circ) \subseteq [p^1, p^2] \cup [p^2, p^3] \cup [p^1, p^3]$, then there exists i, j such that $p(\Delta_2^\circ) \subseteq [p^i, p^j]$*

To state the lemma in words, if the simplex is mapped to the image of the boundary, its image is contained in one edge's image.

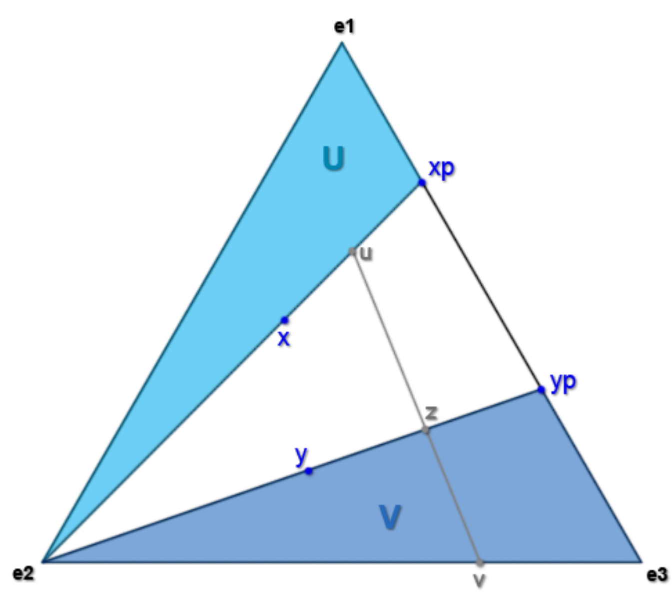
[5] i.e. $p(b_1) \neq p(b_2)$ and $p(a_1) \neq p(a_2)$, and at least one of: $p(b_1) \notin L(p(a_1), p(a_2))$ or $p(b_2) \notin L(p(a_1), p(a_2))$

Proof. If not, then there exist $x, y \in \Delta_2^\circ$ that are mapped to the images of different faces of Δ_2. Without loss of generality $p(x) \in [p^1, p^2] \subset L_{1,2}$, $p(y) \in [p^2, p^3] \subset L_{2,3}$. Because they are assumed to be mapped to different faces $p(x), p(y) \neq p^2$, since if $p(x) = p^2$ we get that $p(x) \in L_{2,3}$, and likewise for $p(y)$.

Let x_p be the point on the intersection of the line $L(e_2, x)$ and $[e_1, e_3]$ (See Fig.1). Because $p^2 \neq p(x)$ any point on $[e_2, x_p] \subset L(e_2, x)$ will be mapped to $L(p^2, p(x)) = L_{1,2}$. Since $p(x) \notin L_{2,3}$ one of $p([e_2, x]), p([x, x_p])$ must have an empty intersection with $L_{2,3}$, as otherwise we get a contradiction to convexity. Let us assume that $p([x, x_p]) \cap L_{2,3} = \phi$ (The case $p([e_2, x]) \cap L_{2,3} = \phi$ is dealt in a similar way).

Denote by $U = conv\{e_2, e_1, x_p\}$ the region bounded between the line $L(e_2, x)$ and the edge $[e_2, e_1]$ (See Fig.1). Denote by $V = conv\{e_2, e_3, y_p\}$ the region bounded between the line $L(e_2, y)$ and the edge $[e_2, e_3]$ (See Fig.1). By convexity every point in U is mapped to the line $L_{1,2}$ and every point in V is mapped to the line $L_{2,3}$.

Figure 1:

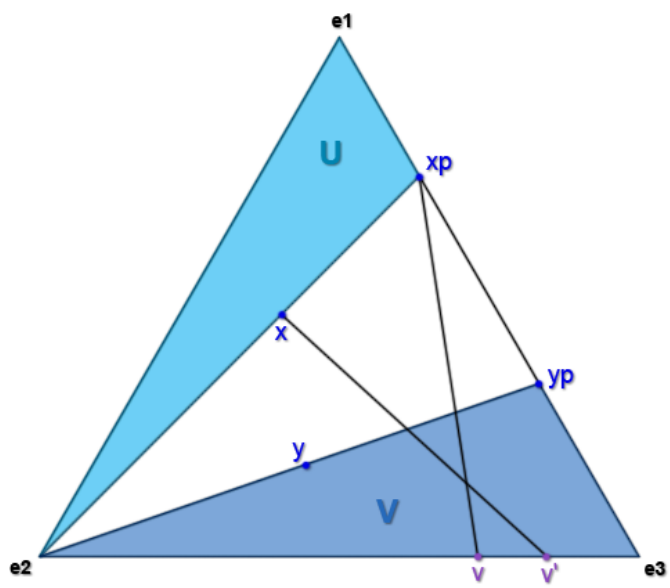


Take two points $v \in (e_2, e_3] \subset V$, $u \in [x, x_p] \subset U$ and consider the segment $[u, v]$.

We can denote $[u, v] \cap V = [z, v]$ for some $z \in V$. Observe that $p(u) \notin L_{2,3}$ and $p([z, v]) \subset L_{2,3}$. By corollary 6 p is constant on $[z, v]$.

Taking $v, v' \in V$, if there are $u, u' \in [x, x_p]$ such that the segments $[v, u]$ and $[v', u']$ intersect at a point $g \in V$ [6], then $p(v) = p(g) = p(v')$ (See Fig.2). Since when $v, v' \in (e_2, e_3]$ are sufficiently close such $u, u' \in [x, x_p]$ can be found, p is locally constant on $(e_2, e_3]$. Any sub-segment $[a, e_3] \subset (e_2, e_3]$ is compact and connected, so p must be constant on $[a, e_3]$. This implies that p is constant on $(e_2, e_3]$, or $p((e_2, e_3]) = \{p^3\}$. Symmetrically $p((e_2, e_1]) = \{p^1\}$. As $\Delta_2^\circ \subseteq conv((e_2, e_1] \cup (e_2, e_3])$ the interior is mapped to the line $[p^1, p^3]$.

Figure 2:



□

Lemma 8 guaranties that when (i) holds and $p(\Delta_2^\circ)$ is not contained in any line then $p(\Delta_2^\circ)$ is not mapped to the boundary of Δ_2. We proceed to prove that (i)

[6] i.e. $g \in [v, u] \cap [v', u'] \cap V$

implies (ii) for the case $n = 3$.

Proposition 9. *The Theorem is true for $n = 3$.*

Proof. We follow the proof of Billot et al. (2005) with minor adaptations. They use a stronger notion of convexity, assuming that $x \in (a, b) \Rightarrow p(x) \in (p(a), p(b)))$. However, the key to their proof is the following condition: if two convex functions, p and p', are identical on four points $\{a_1, a_2, b_1, b_2\}$ – then they are identical on the intersection of the lines $L(a_1, a_2)$,$L(b_1, b_2)$. Indeed, Corollary 7 proves that our convexity suffices for this property, hence the general approach of Billot et al. (2005) would hold in our case as well. Taking into account several additional details, we adapt the Billot et al. (2005) proof for our case.

Under our assumptions there exists a point $b \in \Delta_2^\circ$ that p does not map to the boundary $p(b) \notin [p^1, p^2] \cup [p^2, p^3] \cup [p^1, p^3]$, because otherwise the interior is mapped to a line by the preceding lemma. Observe that for arbitrary $b \in \Delta_2^\circ$, if we start with the points e_1, e_2, e_3, b the set of all points reached by intersection of lines is a dense set. If we use corollary 7 starting with the points e_1, e_2, e_3, b we also obtain that $p(x) = \dfrac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ on a dense subset of Δ_2. Using Billot et al. (2005) completion step we conclude that $p(x) = \dfrac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ on all Δ_2. □

A.2 The induction step

In order to prove the general case we proceed by induction. First, let us find a face whose interior is not mapped to any line. By the induction hypothesis we will get the desired representation on this face. Next we will extend the representation to another face, and determine the values of p on the interior by the values on the two faces. Lastly, we will distinguish between points on the boundary for which the value of p is determined by its value on the interior, and points for which p is not constrained by its values on the interior.

Proposition 10. *Assume $n > 3$. If p is convex and $p(\Delta_{n-1}^\circ)$ isn't contained in any line, then there is a face A_i of Δ_{n-1} such that $p(A_i^\circ)$ is not contained in any line.*

We will use the following lemma for the proof:

Lemma 11. *If $p(\Delta_k^\circ) \subset L$ for some line L and there is a point on the interior of one of the faces of Δ_k, $x \in A_i^\circ$, $A_i \subset \Delta_k$, such that $p(x) \notin L$ then $p(\Delta_k^\circ) = \{c\}$.*

Proof. First, let us show that if there is such a point x then there is also an set $V \subset A_i^\circ$ open relative to A_i, such that $p(v) \notin L, \forall v \in V$. Since L is a convex set, by proposition 4 $p^{-1}(L)$ limited to the face A_i is a convex set. If $p^{-1}(L)$ would have been a dense set in A_i it would follow from the convexity of $p^{-1}(L)$ that it contains the entire interior of A_i. But $x \in A_i^\circ, x \notin p^{-1}(L)$. Hence $p^{-1}(L)$ is not dense, and there exists an open set V such that $V \cap p^{-1}(L) = \phi$.

Now for any two points $a, b \in \Delta_k^\circ$ we can find $\alpha, \beta \in V$ such that $(\alpha, a]$ and $(\beta, b]$ intersect inside Δ_k°. As $p(\alpha) \notin L$ we get by the corollary 6 that $p((\alpha, a]) = \{c_1\}$ and by the same reason $p((\beta, b]) = \{c_2\}$. As the lines intersect $c_1 = c_2$, and we get that $p(a) = p(b)$. As a, b were arbitrary we got that $p(\Delta_k^\circ) = c$ is constant. □

We now turn to prove the proposition:

Proof. We assume the contrary and reach a contradiction. We first show that if the interior of every face is mapped to a line, the lines must all be distinct and non degenerate. This will imply that the interiors of the intersections of faces are mapped to single points. We then identify the values of p on the boundary of the faces and on the vertexes. The resulting structure is impossible when there are more than 3 vertexes, leading to a contradiction that proves our claim.

Let there be a convex p such that $p(\Delta_{n-1}^\circ)$ isn't contained in any line. Let us assume, contrary to our claim, that for every face A_i $p(A_i^\circ) \subset L_i$ for some line L_i. Because Δ_{n-1}° is inside the convex hull of the interior of two faces A_i°, A_j°, their corresponding lines must be different $L_i \neq L_j$ (otherwise the same line $L_i = L_j$ would also contain $p(\Delta_{n-1}^\circ)$). Because Δ_{n-1}° is in the convex hull of the interior of a face and the opposite vertex, p cannot map the interior of a face A_i° to a single point (otherwise the line $L(c, p^i)$, where $p(A_i^\circ) = \{c\}$, would also contain $p(\Delta_{n-1}^\circ)$).

Consider the face A_i as a simplex. If A_i had an edge $A_{i,j}^{\circ} = conv^{\circ}(\{e_1, ..., e_n\} \setminus \{e_i, e_j\})$ that had a point on the interior $x \in A_{i,j}^{\circ}$ such that $p(x) \notin L_i$, by lemma 11 A_i° would be mapped to a point. Therefore for every face A_i all the edges' interiors are also mapped to the line L_i. Explicitly $p(A_{i,j}^{\circ}) \subset L_i$. As the lines are distinct $p((A_i \cap A_j)^{\circ}) \subset L_i \cap L_j = \{c_{i,j}\}$.

Consider a face A_i with interior mapped to L_i. For each of its edges, its interior is mapped to a single point. If there is a vertex $e_j \in A_i$ such that $p(e_j) \notin L_i$, for every $a \in A_i^{\circ}$ consider the segment $[e_j, \hat{a}]$ containing a, when $\hat{a} \in A_{i,j}^{\circ} = conv^{\circ}(\{e_1, ..., e_n\} \setminus \{e_i, e_j\})$ is on the edge opposite to e_j in A_i. By lemma 6 p is constant on $(e_j, \hat{a}]$. Because p is constant on $A_{i,j}^{\circ}$ we get that p is constant on A_i°. Since that would imply $p(\Delta_{n-1}^{\circ})$ is contained in a line, in contradiction to our assumption, we must have for all vertexes $p(e_j) \in L_i$ for every $j \neq i$.

Therefore every vertex e_i is mapped to the point p^i, which is on all the lines L_j for $j \neq i$ and not on L_i, i.e. $\forall i [\forall j \neq i \; p^i \in L_j \, , \, p^i \notin L_i]$. Since the lines $L_j, j = 1 \dots n$ are distinct and $n \geq 4$, this leads to immediate contradiction. □

By proposition 10 we have a face A_i on which we can use the induction hypothesis[7] to get $s_i \in \mathbb{R}_+$ and $p^i \in \Delta_m$ such that $\{p^i | s_i > 0\}$ are not all collinear, and for every $x \in A_i$ for which $\Sigma s_i x_i > 0$ the function p can be put as $p(x) = \dfrac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$. By the following lemma we can extend this representation to another face.

Lemma 12. *If* $p(x) = \dfrac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ *on a face* $A_n = conv(\{e_1, ..., e_{n-1}\})$ *where* $\{p^i | s_i > 0\}$ *are not all collinear, and* $p(\Delta_{n-1}^{\circ})$ *is not contained in any line, then we can find a face* $A_j = conv(\{e_i\}_{i \neq j})$ *and take* s_n *so that* $p(x) = \dfrac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ *on* A_j°*.*

Proof. If there exists a face A_j for which $p(A_j^{\circ})$ is not contained in any line, then there are $\{s_i'\}_{i \neq j}$ so that $p(x) = \dfrac{\Sigma s_i' x_i p^i}{\Sigma s_i' x_i}$ on A_j by the induction step. Notice that for

[7] Note that a face of a simplex is also a simplex

$x \in (A_n \cap A_j)^\circ$ all the coordinates but two are strictly positive. Since at least 3 of the $\{s_i\}$ are strictly positive $\Sigma s_i x_i > 0$, and $\Sigma s_i' x_i > 0$. Therefore, $\frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i} = p(x) = \frac{\Sigma s_i' x_i p^i}{\Sigma s_i' x_i}$ on $(A_n \cap A_j)^\circ$ and we find that $\{s_i'\}_{i \neq j,n}$ are proportional to $\{s_i\}_{i \neq j,n}$. After normalizing $\{s_i'\}$ we can set $s_n = s_n'$ and conclude that $p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ on $A_n^\circ \cup A_j^\circ$.

If the interior of every other face A_i is mapped to a line L_i, by the same manner as in the proof of proposition 10 we find that $\forall i < n\ [\forall j < n, j \neq i\ \ p^i \in L_j, p^i \notin L_i]$. If $n > 4$ this leads to a contradiction.

In the case $n = 4$ we will set $s_4 = 0$ and prove that this extends the representation of p to another face. Because p^1, p^2, p^3 are distinct points $L_i = span\{\{p^1, p^2, p^3\} \setminus \{p^i\}\}$. Since p^1, p^2, p^3 do not lie on one line, the lines L_1, L_2, L_3 are distinct and $L_1 \cap L_2 \cap L_3 = \phi$. Therefore, there exists $i \in \{1, 2, 3\}$ so that $p^4 \notin L_i$. Using corollary 6 we find that all the segments $(e_4, x]$ for $x \in (A_4 \cap A_i)^\circ$ are mapped to a constant. Using the known representation on $(A_4 \cap A_i)^\circ$ and setting $s_4 = 0$ we find that $p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ on A_i°. □

We now have the desired representation on two faces, and by the following proposition we will extend it also to the interior.

Proposition 13. *Let there be a convex mapping p such that $p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ for $x \in A^\circ \cup B^\circ$ for two faces A, B of Δ_{n-1} and $p(A^\circ)$ is not contained in any line, then $p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ on Δ_{n-1}°.*

Proof. Let $y \in \Delta_{n-1}^\circ$. Because $\Delta_{n-1}^\circ \subseteq conv(A^\circ \cup B^\circ)$ there exists $a \in A^\circ, b \in B^\circ$ so that $y \in [a, b]$. We can find a neighborhood $V \subset A^\circ$ of a and a neighborhood $U \subset B^\circ$ of b, such that for every $a' \in V$ there is $b' \in U$ satisfying $y \in [a', b']$. Because $p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$ on A° and $p(A^\circ)$ is not contained in any line, we can find $a' \in V$ such that $p(a') \notin [p(a), p(b)]$.

Assume first that $p(a) = p(b)$. By convexity $p(y) = p(a)$, and

$$\begin{aligned}
\frac{\Sigma s_i y_i p^i}{\Sigma s_i y_i} &= \frac{\Sigma s_i[\lambda a_i + (1-\lambda) b_i] p^i}{\Sigma s_i[\lambda a_i + (1-\lambda) b_i]} = \frac{\lambda \Sigma s_i a_i p^i + (1-\lambda)\Sigma s_i b_i p^i}{\Sigma s_i[\lambda a_i + (1-\lambda) b_i]} \\
&= \frac{\lambda[\Sigma s_i a_i] p(a) + (1-\lambda)[\Sigma s_i b_i] p(b)}{\Sigma s_i[\lambda a_i + (1-\lambda) b_i]} \\
&= \frac{\lambda[\Sigma s_i a_i] p(a) + (1-\lambda)[\Sigma s_i b_i] p(a)}{\Sigma s_i[\lambda a_i + (1-\lambda) b_i]} \\
&= p(a)\frac{\lambda[\Sigma s_i a_i] + (1-\lambda)[\Sigma s_i b_i]}{\Sigma s_i[\lambda a_i + (1-\lambda) b_i]} = p(a)
\end{aligned}$$

Therefore $\frac{\Sigma s_i y_i p^i}{\Sigma s_i y_i} = p(y)$ and the conclusion holds. A similar argument holds if $p(a') = p(b')$.

If $p(a) \neq p(b)$ and $p(a') \neq p(b')$ we find that $L(p(a), p(b)) \neq L(p(a'), p(b'))$ are two distinct well defined lines. We find that $y \in [a, b] \cap [a', b']$, and by corollary 7 we obtain that $p(y) = \frac{\Sigma s_i y_i p^i}{\Sigma s_i y_i}$. □

To finish the proof we show that the result holds for $y \in \partial\Delta_{n-1}$.

Proposition 14. *Let there be a convex mapping* $p : \Delta_{n-1} \to \Delta_m$ *such that* $\forall x \in \Delta^\circ_{n-1} \quad p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$, *then the equality* $p(y) = \frac{\Sigma s_i y_i p^i}{\Sigma s_i y_i}$ *holds for every* $y \in \partial\Delta_{n-1}$ *satisfying* $\Sigma s_i y_i > 0$

Proof. We would like to use the intersecting lines corollary to determine the values of points on the boundary. To fulfill to requirement of the corollary we need to find two lines that meet at y, none of which is mapped to a point.

Let there be a point on the boundary $y \in \partial\Delta_{n-1}$, we will try to determine whether there exists some $x \in \Delta^\circ_{n-1}$ such that $p([x, y)) \neq const$. Taking a parametrization $[x, y) = \{(1-\lambda)x + \lambda y \mid 0 \leq \lambda < 1\}$ we ask if $p(\lambda x + (1-\lambda)y)$ is constant with respect to λ.

We know that for $0 \leq \lambda < 1$ the point $\lambda x + (1-\lambda)y$ is in the interior, hence $p(\lambda x + (1-\lambda)y) = \frac{\Sigma s_i(\lambda x + (1-\lambda)y)_i p^i}{\Sigma s_i(\lambda x + (1-\lambda)y)_i}$. We will take the derivative of this expression with respect to λ and compare it to zero:

$$\frac{\partial p(\lambda x+(1-\lambda)y)}{\partial \lambda} = \frac{\partial}{\partial \lambda}\frac{\Sigma s_i(\lambda x_i+(1-\lambda)y_i)p^i}{\Sigma s_i(\lambda x_i+(1-\lambda)y_i)}$$

Let us denote $P = \Sigma s_i x_i \geq 0, Q = \Sigma s_i y_i \geq 0$ and consider only the numerator of the derivative:

$$\begin{aligned} &= \Sigma s_i p^i(x_i - y_i)\cdot[\lambda P+(1-\lambda)Q] - \Sigma s_i p^i(\lambda x_i+(1-\lambda)y_i)\cdot[P-Q] \\ &= \Sigma s_i p^i[Qx_i - Py_i] \end{aligned}$$

Observe that $P > 0$, and by our assumption $Q = \Sigma s_i y_i > 0$. Denote by B_y the set of internal points x for which $p([x,y)) = const$. If $Q > 0$, the set B_y is the set of solutions to the non-trivial linear constraints $\Sigma s_i p^i[Qx_i - Py_i] = 0$. Hence, if $Q > 0$ the set B_y is of measure zero in Δ°_{n-1}. Therefore, we can find $x_1 \notin B_y$, or in other words, we can find $x_1 \in \Delta^\circ_{n-1}$ such that $p([x_1,y)) \neq const$. Denote by L_{x_1} the line created by $p([x_1,y)$. Observe that the set $\{x \in \Delta^\circ_{n-1} | p(x) \in L_{x_1}\}$ is also of measure zero. Therefore, we can find another point x_2 such that $x_2 \notin B_y$ and $p(x_2) \notin L_{x_1}$. Given the two segments $[x_1,y),[x_2,y)$, which are mapped to distinct lines, we can use the intersecting lines corollary 7 to get that $p(y) = \frac{\Sigma s_i y_i p^i}{\Sigma s_i y_i}$. □

If $Q = 0$, for every i either $y_i = 0$ or $s_i = 0$. We find that $\Sigma s_i p_i[Qx_i - Py_i] = -P\Sigma s_i p_i y_i = 0$. Since this last sum does not depend on x, every line $[x,y)$ is mapped by p to a constant. Thus $p(y)$ cannot be determined by the values of p on Δ°_{n-1}.

A.3 Rational Domain

As stated in the previous sections, a function $p : C^* \to \Delta_m$ that satisfies the invariance axiom and the weak concatenation axiom is only dependent on the relative frequencies of the cases, and can be equivalently represented as a function $p : \Delta_{n-1} \cap \mathbb{Q} \to \Delta_m$ where the vectors in $\Delta_{n-1} \cap \mathbb{Q}$ represent the relative frequencies of the cases. The equivalent of the weak concatenation axiom in terms of mappings of relative frequencies is the following:

Definition. A function $p : \Delta_{n-1} \cap \mathbb{Q} \to \Delta_m$ will be called *convex* if for every $x, a, b \in \Delta_{n-1} \cap \mathbb{Q}$ such that $x \in [a, b]$, p satisfies

$$p(x) \in [p(a), p(b)]$$

The following theorem is equivalent to our main theorem 1:

Theorem 15. *Let there be given a function $p : \Delta_{n-1} \cap \mathbb{Q} \to \Delta_m$. The following are equivalent:*

(i) p is convex and $p((\Delta_{n-1} \cap \mathbb{Q})^\circ)$ is not contained in any line.

(ii) There exist vectors $p^i \in \Delta_m$ and non-negative numbers $\{s_i\}_{i=1}^n$ such that $\{p^i | s_i > 0\}$ are not all collinear, and for every $x \in \Delta_{n-1}$ for which $\Sigma s_i x_i > 0$ the function p can be put as

$$p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$$

and p restricted to $\{x | \Sigma s_i x_i = 0\}$ is convex.

Proof. As stated before, all $\Delta_{n-1} \cap \mathbb{Q} = conv_{\mathbb{Q}}(\{e_1, ..., e_n\})$ is dense in Δ_{n-1} and closed under line intersection. Therefore, all the claims and arguments used for the real case are also valid here.

We should note that some of the proofs use open sets and measure zero sets. There are subtle details here that should be noted. When we refer to a set in $\Delta_{n-1} \cap \mathbb{Q}$ as being open it is meant to be open relative to the subspace topology, for example $(a, b) \cap \mathbb{Q}$ is an open set in $\Delta_{n-1} \cap \mathbb{Q}$ and $(\Delta_{n-1} \cap \mathbb{Q})^\circ = \{\sum \alpha_i e_i | \sum \alpha_i = 1, \alpha_i \in \mathbb{Q}, \alpha_i > 0\}$.

In the proof of proposition 14 we use the notion of measure zero sets. As the whole set $\Delta_{n-1} \cap \mathbb{Q}$ is of measure zero we need to refine the argument in the proof. Instead using the notion of measure zero sets we can argue that the sets mentioned in the proof have a closer that is of measure zero and therefore cannot contain the entire $\Delta_{n-1} \cap \mathbb{Q}$. The rest of the proof remains the same. $\square$

B Lexicographic Similarity

Let us look at those points whose values the theorem did not determine. Denote this set by $D_1 = conv\{e_i | s_i = 0\}$. For $y \in D_1$ we found that $Q = \Sigma s_i y_i = 0$, and by the preceding comment $p([x, y)) = const$ for all $x \notin D_1$. Therefore the values of p on D_1 can be arbitrarily changed without affecting the convexity of p, as long as p is convex on D_1.

We have that $p : D_1 \to \Delta_m$ is a function which does not depend on $p|_{\Delta_{n-1} \setminus D_1}$. Because p is convex function, if $p(D_1^\circ)$ is not contained in a line we can use the theorem to obtain

$$p(x) = \frac{\Sigma s_i^1 x_i p^i}{\Sigma s_i^1 x_i}$$

for $x \in D_1$ such that $\Sigma s_i^1 x_i > 0$.

This process can be followed iteratively, producing weights $\{s_i^2\}$ for p restricted to $D_2 = conv\{e_i | s_i^1 = 0\}$, and producing at the k-th step weights $\{s_i^k\}$ for p restricted to $D_n = conv\{e_i | s_i^{k-1} = 0\}$. This process will end only when we reach a k such that $p(D_k^\circ) \subset L$ for some line L.

By the process we obtained an hierarchy of weights $\{\{s_i^k\}\}_{k=1}^K$. For $x \in \Delta_{n-1}$ and the smallest k for which $\Sigma s_i^k x_i > 0$ the equality $p(x) = \dfrac{\Sigma s_i^k x_i p^i}{\Sigma s_i^k x_i}$ holds. No such k can be found only for $x \in D_K$.

This hierarchy of weights can be given a more concise "numerical" representation by giving the weights values from a non-Archimedean ordered field $\mathbb{F}$ which is a strict extension of the real numbers fields $\mathbb{R}$. The field $\mathbb{F}$ is non-Archimedean: it contains, in addition to the reals, both infinite numbers (larger than any real number) and infinitesimal numbers (positive, but smaller than any positive real number) .

Let $\varepsilon \in \mathbb{F}$ be an infinitesimal . We can chose the weights s_i to be

$$s_i = \varepsilon^k s_i^k \ , \ k = min\{k | s_i^k \neq 0\}$$

allowing us to state that for every x such that $\Sigma s_i x_i > 0$

$$p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$$

describing p on all $\Delta_{n-1} \cap \mathbb{Q} \setminus D_K$.

We have thus proved the following version of our main theorem 2:

Theorem 16. *Let there be given a function* $p : \Delta_{n-1} \cap \mathbb{Q} \to \Delta_m$*. The following are equivalent:*

(i) p *is convex.*

(ii) There exists vectors $p^i \in \Delta_m$ *and non-negative weights* $\{s_i\}_{i=1}^n$ *taken from a non-Archimedean ordered field* $\mathbb{F}$*, such that for every* $x \in \Delta_{n-1} \cap \mathbb{Q}$ *for which* $\Sigma s_i x_i > 0$ *the function* p *can be put as*

$$p(x) = \frac{\Sigma s_i x_i p^i}{\Sigma s_i x_i}$$

In addition p *restricted to* $D = \{x | \Sigma s_i x_i = 0\}$ *is convex, and* $p(D^\circ) \subset L$ *for some line* L*.*