Probabilistic Merging Operators

Adamcik, Martin and Wilmers, George

2013

MIMS EPrint: **2013.41**

ISSN 1749-9097

Probabilistic Merging Operators

Martin Adamčík* **and George Wilmers**
School of Mathematics
The University of Manchester
martin.adamcik@manchester.ac.uk, george.wilmers@gmail.com

30th April 2013

Abstract

The present work presents a general theoretical framework for the study of operators which merge *partial* probabilistic evidence from different sources which are individually coherent, but may be collectively incoherent. We consider a number of principles for such an operator to satisfy including a set of principles derived from those of Konieczny and Pino Pérez [11] which were formulated for the different context of propositional merging. Finally we investigate two specific such merging operators derived from the Kullback-Leibler notion of informational distance: the social entropy operator, and its dual, the linear entropy operator. The first of these is strongly related to both the multi-agent normalised geometric mean pooling operator and the single agent maximum entropy inference process, **ME**. By contrast the linear entropy operator is similarly related to both the arithmetic mean pooling operator and the limit centre of mass inference process, $\mathbf{CM}^{\infty}$.

Keywords: uncertain reasoning, probability function, merging of evidence, Kullback-Leibler, divergence, probabilistic merging, merging operator, Konieczny and Pino Pérez, social entropy process, inference process, aggregation of probabilities, pooling operator, probabilistic inference, maximum entropy.

1 Introduction

This work studies some of the global logical desiderata which a well-defined process for merging partial probabilistic evidence should satisfy. The probabilistic evidence is thought of as arising from a set sources each of which provides coherent probabilistic evidence, while the collective evidence from all the sources is typically inconsistent. The objective of such a merging process is to merge

*The first author was supported by a Marie Curie Early Stage Researcher Fellowship (MALOA - from Mathematical Logic to Applications) financed by the European Commission.

the evidence of a set of such sources into a single coherent evidence base, which best represents the declared evidence of all the sources, but *no more*.

The result of such a merging process will not always be a single probability function, but rather a non-empty closed convex set of such functions which *represents the merged evidence* as if from the standpoint of an unbiased external observer with no evidence of her own. If for pragmatic purposes a single probability function must be chosen, then this choice can be made at a second stage by the use of whatever single-agent inference process is preferred.

Much axiomatic analysis has been done previously on the very special case of probabilistic pooling or aggregation operators[1], where each individual source offers a single probability distribution as evidence. Furthermore many different algorithms have been suggested for dealing with the far more complex case of probabilistic merging where each source offers only partial probabilistic evidence[2]. However few authors have considered the global desiderata which such general probabilistic merging should satisfy, and where such desiderata have been considered, many authors have not clearly distinguished the operation of merging the evidence bases from the goal of choosing a *unique* probability function from the merged evidence. One exception in this respect is Williamson who stresses the philosophical distinction between these two processes, and has sought to adapt[3] to the probabilistic context the norms for propositional merging which were first formulated by Konieczny and Pino Pérez in [11].

We believe that Williamson's distinction above is a useful one. In this paper we formulate a probabilistic adaptation of the Konieczny and Pino Pérez principles. We then investigate in this context the properties of two particular probabilistic merging operators, social entropy, and linear entropy, which are respectively generalisations of the normalised geometric mean and linear pooling operators. Social entropy was defined in [22], and was shown in [23] to bear a natural relationship to the well-known maximum entropy inference process $\mathbf{ME}$[4]. On the other hand linear entropy, which is a dual merging operator to social entropy, bears a corresponding natural relationship to $\mathbf{CM}^{\infty}$, the limit centre of mass inference process[5].

Konieczny and Pino Pérez in [11] proposed an axiomatic framework, referred to below as KPP, for expressing the desiderata required of a *propositional* merging operator. Such an operator Δ acts on a multiset of knowledge bases $T_1, \ldots, T_n$ to generate a single knowledge base. Each knowledge base T_i is assumed consistent[6], but the union of two or more knowledge bases may not be

[1] See e.g. [9] for a survey
[2] See e.g. [10],[14],[18],[22],[20],[21].
[3] See [19],[20],[21].
[4] See [15] or [17] for a detailed characterisation of $\mathbf{ME}$.
[5] See [15] for a definition of $\mathbf{CM}^{\infty}$.
[6] This is a slight restriction of the KPP formulation which is more appropriate for our

consistent. The resulting *merged* knowledge base $\Delta(T_1, \ldots, T_n)$ should be consistent, and the operator Δ should at a minimum satisfy the principles listed below. In [11] the case was considered where a knowledge base is interpreted to mean a consistent set of sentences of a given finite propositional language L. However, as noted in [11], the general idea of a merging operator can easily be applied to other types of knowledge base, and there exists a large literature concerning such generalisations[7].

In the KPP framework described above a merging operator Δ should satisfy the following principles:

For every $n, m \geq 1$ and every propositional language L and knowledge bases $K_1, \ldots, K_n, F_1, \ldots, F_m$ for L:

(A1) $\Delta(K_1, \ldots, K_n)$ is a knowledge base,

(A2) If $K_1, \ldots, K_n$ are jointly consistent then $\Delta(K_1, \ldots, K_n)$ is logically equivalent to $\bigcup_{i=1}^{n} K_i$,

(A3) If $K_1, \ldots, K_n$ and $F_1, \ldots, F_n$ are such that there exist a permutation π of the index set $\{1, \ldots, n\}$ such that K_i is logically equivalent to $F_{\pi(i)}$ for $1 \leq i \leq n$, then $\Delta(K_1, \ldots, K_n)$ is logically equivalent to $\Delta(F_1, \ldots, F_n)$,

(A4) If K_1 and F_1 are jointly inconsistent then $\Delta(K_1, F_1) \not\models K_1$,

(A5) $\Delta(K_1, \ldots, K_n) \cup \Delta(F_1, \ldots, F_m) \models \Delta(K_1, \ldots, K_n, F_1, \ldots, F_m)$,

(A6) If $\Delta(K_1, \ldots, K_n) \cup \Delta(F_1, \ldots, F_m)$ is consistent then

$$\Delta(K_1, \ldots, K_n, F_1, \ldots, F_m) \models \Delta(K_1, \ldots, K_n) \cup \Delta(F_1, \ldots, F_m).$$

In the next section we will reformulate the ideas behind the KPP principles above in order to apply them to the different context of the merging of *probabilistic evidence bases*, or more explicitly, to the search for an objective method of merging probabilistic evidence from distinct sources into a single coherent evidence base.

Before continuing our discussion we will now formulate the prerequisite concepts which we will need in order to define precisely the general notion of a probabilistic merging operator.

2 From Propositional to Probabilistic Merging

Let $L = \{p_1 \ldots p_h\}$ be a finite propositional language where $p_1, \ldots, p_h$ are propositional variables. We denote the set of all propositional sentences which are

present considerations.

[7] See [12] for a survey paper and bibliography.

possible to define over L as SL. By the disjunctive normal form theorem any L-sentence can be expressed as a disjunction of atomic sentences (atoms), and we will denote by $\mathrm{At}(L)$ some fixed maximal set of logically inequivalent atoms $\{\alpha_1, \ldots, \alpha_J\}$, where $J = 2^h$. The atoms of $\mathrm{At}(L)$ are thus mutually exclusive and exhaustive.

A *probability function* $\mathbf{w}$ over L is defined as a function $\mathbf{w} : \mathrm{At}(L) \to [0,1]$ such that $\sum_{j=1}^{J} \mathbf{w}(\alpha_j) = 1$. A value of $\mathbf{w}$ on any L-sentence φ may then be defined by setting

$$\mathbf{w}(\varphi) = \sum_{\alpha_j \models \varphi} \mathbf{w}(\alpha_j).$$

We will denote the set of all probability functions over L by $\mathbb{D}^L$. For the sake of simplicity we will often write w_j instead of $\mathbf{w}(\alpha_j)$, but note that this makes sense only for atoms. Given a probability function $\mathbf{w} \in \mathbb{D}^L$, a conditional probability is defined by Bayes's formula

$$\mathbf{w}(\varphi|\psi) = \frac{\mathbf{w}(\varphi \wedge \psi)}{\mathbf{w}(\psi)}$$

for any L-sentence φ and any L-sentence ψ such that $\mathbf{w}(\psi) \neq 0$ and is left undefined otherwise.

A *probabilistic evidence base* $\mathbf{K}$ over L is a set of constraints on probability functions over L such that the set of all probability functions satisfying the constraints in $\mathbf{K}$ forms a nonempty closed convex subset $V^L_{\mathbf{K}}$ of $\mathbb{D}^L$. For brevity we shall use the terminology *evidence base* instead of probabilistic evidence base. $V^L_{\mathbf{K}}$ may be thought of as the set of possible probability functions in $\mathbb{D}^L$ of a particular agent which are consistent with her evidence base $\mathbf{K}$. We shall generally write $V_{\mathbf{K}}$ instead of $V^L_{\mathbf{K}}$ unless there is any ambiguity about which language is referred to. Note that this standard formulation ensures that linear constraint conditions such as $\mathbf{w}(\theta) = a$, $\mathbf{w}(\phi \mid \psi) = b$, and $\mathbf{w}(\psi \mid \theta) \leq c$, where $a, b, c \in [0,1]$ and θ, ϕ, and ψ are L-sentences, are all permissible in an evidence base $\mathbf{K}$ provided that the resulting constraint set $\mathbf{K}$ is consistent with the laws of probability. Note that a constraint such as $\mathbf{w}(\psi \mid \theta) \leq c$ is interpreted as $\mathbf{w}(\psi \wedge \theta) \leq c \cdot \mathbf{w}(\theta)$ which makes sense as a linear constraint even though $\mathbf{w}(\theta)$ may take the value zero (see [15] for details).

If $\mathbf{K}_1$ and $\mathbf{K}_2$ are such that $V_{\mathbf{K}_1} = V_{\mathbf{K}_2}$ we shall say that $\mathbf{K}_1$ and $\mathbf{K}_2$ are *equivalent*. In practice we shall only be interested in constraint sets up to equivalence, and consequently we may sometimes informally identify an evidence base $\mathbf{K}$ with its extension $V_{\mathbf{K}}$, and with slight abuse of language we may also refer to a non-empty closed subset of $\mathbb{D}^L$ as an evidence base. Note that the non-emptiness of $V_{\mathbf{K}}$ corresponds to the assumption that $\mathbf{K}$ is consistent with the laws of probability, while if $\mathbf{K}$ and $\mathbf{F}$ are evidence bases then the set of constraints $\mathbf{K} \cup \mathbf{F}$ corresponds to $V_{\mathbf{K} \cup \mathbf{F}} = V_{\mathbf{K}} \cap V_{\mathbf{F}}$, and so forms an evidence base provided that the latter intersection is non-empty.

The set of all evidence bases $V_{\mathbf{K}}$ over L is denoted by CL. A more restricted notion of evidence base is an evidence base *which bounds probability functions away from zero*. This is an evidence base $\mathbf{K} \in CL$ such that $V_{\mathbf{K}}$ satisfies a set of constraints on $\mathbf{w} \in V_{\mathbf{K}}$ of the form

$$\{a_j \leq w_j\colon\ 1 \leq j \leq J\}$$

where $0 < a_j < 1$ for all $j = 1 \ldots J$. We call such an evidence base *bounded*, and we will denote the set of all bounded evidence bases for a given language L by BCL. A slightly more general notion is that of an evidence base $\mathbf{K} \in CL$ which does not "force" any atom to take the value zero. More precisely we call $\mathbf{K}$ *weakly bounded* if for every $1 \leq j \leq J$ there is $\mathbf{w} \in V_{\mathbf{K}}$ such that $w_j \neq 0$. The set of weakly bounded evidence bases for L will be denoted by $WBCL$. Note that $BCL \subset WBCL \subset CL$ and that by convexity if $\mathbf{K} \in WBCL$ then there exists some $\mathbf{w} \in V_{\mathbf{K}}$ such that $w_j \neq 0$ for all $j = 1 \ldots J$.

There are at several possible motivations for studying evidence bases with a boundedness condition imposed. Broadly speaking, the imposition of such a condition may avoid some of the potentially intractable technical and philosophical difficulties which arise from treating zero probabilities in certain contexts. In this paper we will confine ourselves to stating and proving some theorems concerning particular merging operators for certain classes of evidence base, but will not consider further the epistemological status of the various notions of evidence base.

Let Δ denote an operator defined for all $n \geq 1$ and all L as a mapping

$$\Delta_L : \underbrace{CL \times \ldots \times CL}_{n} \to \mathcal{P}(\mathbb{D}^L)$$

where $\mathcal{P}(\mathbb{D}^L)$ denotes the power set of $\mathbb{D}^L$. We will call such a Δ a *probabilistic merging operator*, abbreviated to *p-merging operator*, if it satisfies the following

(K1) Defining Principle.

If $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ then the set $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n) \in CL$.

Note that **(K1)** is a natural counterpart to **(A1)**; just as **(A1)** ensures that a propositional merging operator applied to a multiset of knowledge bases yields a knowledge base, so **(K1)** ensures that a p-merging operator applied to a multiset of evidence bases yields an evidence base.

In general we shall suppress the subscript L in Δ_L except where an ambiguity could be caused by such an omission. We may sometimes slightly abuse the above terminology by referring to an operator Δ as a p-merging operator even though the domain over which Δ is properly defined may be a certain subclass of the $\underbrace{CL \times \ldots \times CL}_{n}$. Whenever we do this however the correct restriction of the domain of application will always be made apparent.

We now set about reformulating the remaining KPP principles so as to make them applicable to the context of a p-merging operator Δ. We express the remaining four principles as follows:

For every $n \geq 1$ and every propositional language L

(K2) Consistency Principle. For all $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ if $\bigcap_{i=1}^n V^L_{\mathbf{K}_i} \neq \emptyset$ then $\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) = \bigcap_{i=1}^n V^L_{\mathbf{K}_i}$.

(K2) can be interpreted as saying that if the evidence bases of a set of agents are collectively consistent then the merged evidence base should simply consist of all the evidence of the agents collected together.

(K3) Equivalence Principle. If $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ and $\mathbf{F}_1, \ldots, \mathbf{F}_n \in CL$ are such that there exist a permutation π of the index set $\{1, \ldots, n\}$ such that $V^L_{\mathbf{K}_i} = V^L_{\mathbf{F}_{\pi(i)}}$ for $1 \leq i \leq n$, then $\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) = \Delta(\mathbf{F}_1, \ldots, \mathbf{F}_n)$.

Notice that **K3** has the effect that for any Δ which satisfies it, the order in which the evidence bases occur when Δ is applied is immaterial, and therefore we can loosely refer to Δ as being applied to a multiset of evidence bases instead of a sequence of such evidence bases. On the other hand repetitions of evidence bases *will* in general be significant, so the sequence (or multiset) of evidence bases cannot be considered as a set; the Δ we consider will typically share most of the characteristics of what are commonly referred to as majority merging operators.

(K4) Disagreement Principle. Let $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ and $\mathbf{F}_1, \ldots, \mathbf{F}_m \in CL$. Assume that $\bigcap_{i=1}^m V^L_{\mathbf{F}_i} \neq \emptyset$.
Then $\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \Delta(\mathbf{F}_1, \ldots, \mathbf{F}_m) = \emptyset$ implies that

$$\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n, \mathbf{F}_1, \ldots, \mathbf{F}_m) \cap \Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) = \emptyset.$$

(K4) represents a significant but natural strengthening of **(A4)**, adapted to the p-merging context. Intuitively the principle says that if the merged evidence base $\mathbf{K}$ of a set of agents is inconsistent with the merged evidence $\mathbf{F}$ of a distinct set of agents, where the evidence bases of the latter set are collectively consistent, then the result of merging the evidence bases of all the agents together is also inconsistent with $\mathbf{K}$. Expressed more pithily, but less exactly, we could say that a coherent group who disagree with another group and then merge with them can be sure that they have influenced the opinions of the combined group.

(K5) Agreement Principle. If $\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \Delta(\mathbf{F}_1, \ldots, \mathbf{F}_m) \neq \emptyset$ then

$$\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \Delta(\mathbf{F}_1, \ldots, \mathbf{F}_m) = \Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n, \mathbf{F}_1, \ldots, \mathbf{F}_m).$$

(K5) combines the ideas of **(A4)** and **(A5)** into a single principle adapted to the probabilistic context. In particular **(K5)** implies that if each of two distinct sets of agents arrive at the same set of possible conclusions then the result of considering the evidence bases of all the agents together should result in the same set of possible conclusions.

The intuitive idea behind p-merging is that the probabilistic evidence from a set of agents should be shared on an equal basis by some objective collaborative process, which takes full account of the declared evidence base of each agent, including the *implicit ignorance* of each agent whenever an agent has not specified a singleton probability function as constituting her evidence base. The result of this process should be a new "social" or merged evidence base, which represents the total declared evidence of the set of agents, just as if the set had merged to form a single agent. As in the case of the evidence base of an individual agent, it is clear from **(K2)** that this merged evidence base will not in general be a singleton. It should be viewed instead as the set of those possible probability functions which are rationally generated by the *total evidence* of the set of agents in the absence of any other information.

This general intersubjective approach to probabilistic merging was expounded in a slightly different form by the second author in [22], and accords well with certain philosophical ideas elaborated independently by Williamson [20],[19]. Both stress the advantages of initially merging the *evidence bases* of a set of agents into a single evidence base, as opposed to merging the probabilistic *beliefs* of the individual agents, i.e. the unique probability functions which each agent may hypothetically arrive at solely by considering her own evidence base and applying to it a standard inference process[8] such as the maximum entropy inference process **ME**.

Our reformulation of the KPP principles into a probabilistic framework is a fairly straightforward translation with the exception, as noted above, of **(K4)**. We should note however that whereas Williamson previously advocated the relevance of the KPP principles in relation to the merging of evidence bases, in a more recent paper [21] he rejects the KPP principles **(A2)**, **(A4)**, and **(A6)** as representing norms which are too strong to be applicable in this context.

[8]See e.g. [15] for a comprehensive account of single agent inference processes, including **ME**.

However in order to arrive at this conclusion Williamson uses a particular interpretation of the epistemological status of an agent's evidence base, which he calls "granting", and which we do not share; for this reason we do not find Williamson's arguments against these principles to be conclusive. Furthermore, as we will show later, the hitherto known p-merging operators which satisfy perhaps the most attractive desiderata other than **(K1)**-**(K5)** do in fact satisfy the above principles **(K1)**, **(K2)**, **(K3)**, and satisfy **(K4)** and **(K5)** at least when their application is restricted to bounded evidence bases.

In [22] and [23] a specific p-merging operator is defined, which we will here call the Social Entropy operator[9], denoted by Δ^{KL}, which is strongly related to Kullback-Leibler divergence. In the next section we will examine in more detail some of the properties of the operator Δ^{KL} together with those of its dual, the Linear Entropy operator $\hat{\Delta}^{\mathrm{KL}}$.

3 Two Probabilistic Merging Operators

3.1 The Social Entropy Operator Δ^{KL}

In order to define the social entropy operator we first need to define Kullback-Leibler divergence KL : $\mathbb{D}^L \times \mathbb{D}^L \to \mathbb{R} \cup \{+\infty\}$. This may be thought of as a function which measures of the (asymmetric) "informational distance" from one probability function to another, and returns a value in the interval $[0, +\infty]$. The asymmetry of this notion is the reason for the use of the term "divergence" rather than "distance". The divergence from $\mathbf{w} \in \mathbb{D}^L$ to $\mathbf{v} \in \mathbb{D}^L$ is $+\infty$ whenever $v_j \neq 0$ and $w_j = 0$ for some atom α_j. If this is not the case we say that $\mathbf{w}$ *dominates* $\mathbf{v}$ and write $\mathbf{w} \gg \mathbf{v}$. Let $\mathrm{Sig}(\mathbf{w}) = \{j\colon\ w_j \neq 0\}$. Then the Kullback-Leibler divergence is defined by

$$\mathrm{KL}(\mathbf{v}\|\mathbf{w}) = \begin{cases} \sum_{j \in \mathrm{Sig}(\mathbf{w})} v_j \log \frac{v_j}{w_j} & \text{if } \mathbf{w} \gg \mathbf{v}, \\ +\infty & \text{otherwise.} \end{cases}$$

with the usual convention that $x \log x$ is defined to take the value zero at $x = 0$.

[9] In [22] and [23] this was introduced as the first stage of a merging operator called the social entropy process, **SEP**, which for any multiset of evidence bases chooses a **unique** merged probability function. The first stage consists of applying Δ^{KL} and the second stage chooses the unique maximum entropy point in the resulting merged evidence base.

Given evidence bases $\mathbf{K}_1,\ldots,\mathbf{K}_n \in CL$ let

$$C_{\mathbf{K}_1,\ldots,\mathbf{K}_n} \;=\; \min\{\sum_{i=1}^{n} \mathrm{KL}(\mathbf{v}\|\mathbf{w}^{(i)}) \,:\, \mathbf{v}\in\mathbb{D}^L\,;\, \mathbf{w}^{(1)}\in V^L_{\mathbf{K}_1},\ldots,\mathbf{w}^{(n)}\in V^L_{\mathbf{K}_n}\}$$

It is easy to see that this is well-defined (see [23]). Note that this value lies in the interval $[0,+\infty]$. Also $C_{\mathbf{K}_1,\ldots,\mathbf{K}_n}=0$ if and only if $\mathbf{v}=\mathbf{w}^{(1)}=\ldots=\mathbf{w}^{(n)}$ in the definition above in which case the $\mathbf{K}_1,\ldots,\mathbf{K}_n$ are jointly consistent. Also $C_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$ is finite if and only if the following holds:

There is some atom α_j such that for no i is it the case that for all $\mathbf{w}\in V^L_{\mathbf{K}_i}$ $\mathbf{w}(\alpha_j)=0$. (1)

The p-merging operator Δ^{KL} is now defined as follows: for any L and any $\mathbf{K}_1,\ldots,\mathbf{K}_n \in CL$ $\Delta^{\mathrm{KL}}_L(\mathbf{K}_1,\ldots,\mathbf{K}_n)$ is defined as

$$\{\mathbf{v}\in\mathbb{D}^L \,:\, \exists\,\mathbf{w}^{(1)}\in V^L_{\mathbf{K}_1},\ldots,\mathbf{w}^{(n)}\in V^L_{\mathbf{K}_n} \text{ s.t. } \sum_{i=1}^{n}\mathrm{KL}(\mathbf{v}\|\mathbf{w}^{(i)}) \;=\; C_{\mathbf{K}_1,\ldots,\mathbf{K}_n}\}$$

In [23] it is shown that for any $\mathbf{K}_1,\ldots,\mathbf{K}_n \in CL$ this set $\Delta^{\mathrm{KL}}_L(\mathbf{K}_1,\ldots,\mathbf{K}_n)$ is always a non-empty closed convex region of $\mathbb{D}^L$, and hence it follows that the p-merging operator Δ^{KL} satisfies **(K1)**. We note however that although Δ^{KL} is everywhere defined[10] it is really only interesting as a merging operator for those $\mathbf{K}_1,\ldots,\mathbf{K}_n \in CL$ for which the relatively undemanding condition (1) above is satisfied, since otherwise applying Δ^{KL} simply returns the whole space $\mathbb{D}^L$. The fact that social entropy operator Δ^{KL} satisfies **(K2)** follows at once from the fact noted above that $C_{\mathbf{K}_1,\ldots,\mathbf{K}_n}=0$ if and only if $\mathbf{v}=\mathbf{w}^{(1)}=\ldots=\mathbf{w}^{(n)}$ in the definition of $C_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$. Moreover Δ^{KL} satisfies **(K3)** trivially by definition.

Δ^{KL} turns out to have many other desirable properties, some of which closely resemble the axiomatic properties which have been used to characterise the **ME** inference process in [17], and [15]. (See [22], [23], [1] for details.) In particular we mention the following:

1. **Language Invariance**

   Suppose $L\subset L'$ and $\mathbf{K}_1,\ldots,\mathbf{K}_n \in CL$. $\mathbf{K}_1,\ldots,\mathbf{K}_n$ may also be regarded as evidence bases in CL'. For any $\mathbf{w}'\in\mathbb{D}^{L'}$ denote by $\mathbf{w}'\restriction L$ the marginalisation of $\mathbf{w}'$ to $\mathbb{D}^L$. Then

$$\Delta^{\mathrm{KL}}_L(\mathbf{K}_1,\ldots,\mathbf{K}_n)=\{\mathbf{w}'\restriction L \,:\, \mathbf{w}'\in\Delta^{\mathrm{KL}}_{L'}(\mathbf{K}_1,\ldots,\mathbf{K}_n)\}$$

[10] In the presentation in [23] the region $\Delta^{\mathrm{KL}}_L(\mathbf{K}_1,\ldots,\mathbf{K}_n)$ is only defined assuming that condition (1) holds, but this does not significantly affect the results.

Language Invariance means that if we change a multiset of evidence bases only by adding propositional variables to the language in which they are formulated but add no new evidence, then the restriction of the new merged evidence base to the original language is the same as the original merged evidence base. The fact that Δ^{KL} satisfies language invariance is proved in [1].

2. **The Consistent Irrelevant Information Principle.**

   Let $L = L_1 \cup L_2$ where L_1 and L_2 are disjoint propositional languages. Let $\mathbf{K}_1, \ldots, \mathbf{K}_n$ and $\mathbf{F}_1, \ldots, \mathbf{F}_n$ be knowledge bases formulated for the languages L_1 and L_2 respectively, and suppose that $\mathbf{F}_1, \ldots, \mathbf{F}_n$ are jointly consistent. Then

$$\Delta_L^{\mathrm{KL}}\left(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n\right) \restriction L_1 = \Delta_L^{\mathrm{KL}}\left(\mathbf{K}_1, \ldots, \mathbf{K}_n\right) \restriction L_1 .$$

   where $\restriction L_1$ denotes marginalisation to $\mathbb{D}^{L_1}$. This property of Δ^{KL} follows from Lemma 5.2 of [1]. Together with language invariance, it ensures that if a set of agents have evidence bases formulated in the language L_1 then their merged evidence base remains the same if each agent acquires additional new evidence formulated in a disjoint language L_2 and the newly merged evidence base of all the agents is then restricted to the language L_1, *provided that* all the new evidence in the language L_2 is jointly consistent.

3. Δ^{KL} **Generalises the LogOp Pooling Operator**

   In [22] the following equivalence between (i) and (ii) below is given, which provides an alternative characterisation of Δ^{KL} in the case when condition (1) above is satified:

   (i) The L-probability functions $\mathbf{v}, \mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}$ minimize $\sum_{i=1}^n \mathrm{KL}(\mathbf{v}\|\mathbf{w}^{(i)})$ subject only to $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$.

   (ii) The L-probability functions $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}$ maximize $\sum_{j=1}^J (\prod_{k=1}^n w_j^{(k)})^{\frac{1}{n}}$, subject only to $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots,\ \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$, and

$$v_j = \frac{(\prod_{k=1}^n w_j^{(k)})^{\frac{1}{n}}}{\sum_{j=1}^J (\prod_{k=1}^n w_j^{(k)})^{\frac{1}{n}}} \text{ for all } j = 1, \ldots, J. \tag{2}$$

   Whenever (2) holds we write $\mathbf{v} = \mathbf{LogOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$. **LogOp** is of course just the normalised geometric mean, or "logarithmic", pooling operator familiar to decision theorists. Thus we see that for $\mathbf{v}$ to be in Δ^{KL} there must exist some $\mathbf{w}^{(i)} \in V_{\mathbf{K}_i}$ which maximise the normalising factor in the definition of logarithmic pooling, and for which $\mathbf{v} = \mathbf{LogOp}(\mathbf{w}^{(1)} \ldots \mathbf{w}^{(n)})$. In the very special case when each each agent i specifies a single probability function $\mathbf{w}^{(i)}$ then

$\Delta_L^{\mathrm{KL}}(\mathbf{K}_1,\ldots,\mathbf{K}_n)$ is just the singleton $\{\mathbf{LogOp}(\mathbf{w}^{(1)},\ldots,\mathbf{w}^{(n)})\}$. Notice that condition (1) is exactly the condition required to ensure that the **LogOp** pooling operator is defined.

4. **Δ^{KL} is a Natural Companion to the ME Inference Process**

   At first sight this assertion might seem strange, since if Δ^{KL} is applied to the evidence base $\mathbf{K}$ of a single agent X it simply returns the same evidence base in the form $V_{\mathbf{K}}$, which does not help X to choose a single preferred point in $V_{\mathbf{K}}$. However let us imagine that X now appoints a fanatically unbiased oracle Y with evidence base $\mathbf{F} = \{< \frac{1}{J}, \frac{1}{J} \ldots \frac{1}{J} >\}$, in order to help her to choose a preferred point in her evidence base. Y advises X to imagine cloning herself n times, for some large n, and forming a committee of $n+1$ members consisting of the n clones of X, together with Y as chairman. Finally Y advises X to compute the result of applying Δ^{KL} to the $n+1$ evidence bases of the members of A_n and then to let $n \to \infty$. The result of this procedure is that the merged evidence bases converge towards a single point, the maximum entropy point of $V_{\mathbf{K}}$. (See [23] for a proof[11].)

The following theorem is our first main result of the present work.

Theorem 3.1. *The p-merging operator* Δ^{KL} *satisfies the principles* **(K1)**, **(K2)** *and* **(K3)**. *Furthermore* Δ^{KL} *satisfies* **(K4)** *and* **(K5)** *provided that the evidence bases to which* Δ^{KL} *is applied are restricted to* $WBCL$.

The fact that **(K1)**, **(K2)** and **(K3)** hold for Δ^{KL} has been established above. The rest of the theorem will be proved in section 4.

□

3.2 The Linear Entropy Operator $\hat{\Delta}^{\mathrm{KL}}$

The Linear Entropy operator $\hat{\Delta}^{\mathrm{KL}}$ is a p-merging operator which may naturally be considered as the dual of the Δ^{KL} p-merging operator defined above.

In brief, whereas $\Delta^{\mathrm{KL}}(\mathbf{K}_1,\ldots,\mathbf{K}_n)$ comprises those $\mathbf{v}$ which globally minimise

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{v}\|\mathbf{w}^{(i)})$$

[11] In [23] a similar more general result is proved which holds for any number of agents.

$\hat{\Delta}^{\mathrm{KL}}\ (\mathbf{K}_1, \ldots, \mathbf{K}_n)$ comprises those $\mathbf{v}$ which globally minimise

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v})$$

Given evidence bases $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ let

$$\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n} \;=\; \min\{\sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v}) \;:\; \mathbf{v} \in \mathbb{D}^L \;;\; \mathbf{w}^{(1)} \in V^L_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V^L_{\mathbf{K}_n}\}$$

As in 3.1 it is easy to see that this is well-defined, non-negative, and zero if and only if $\mathbf{v} = \mathbf{w}^{(1)} = \ldots = \mathbf{w}^{(n)}$ in the definition of $\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$. However unlike the case for Δ^{KL} we may note that $\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$ is always finite since any $\mathbf{v}$ all of whose coordinates are non-zero will always give a finite non-zero value to $\sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v})$.

The p-merging operator $\hat{\Delta}^{\mathrm{KL}}$ is now defined as follows: for any L and any $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL \quad \hat{\Delta}^{\mathrm{KL}}_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is defined as

$$\{\mathbf{v} \in \mathbb{D}^L \;:\; \exists\, \mathbf{w}^{(1)} \in V^L_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V^L_{\mathbf{K}_n} \text{ s.t. } \sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v}) \;=\; \hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n}\}.$$

It is easy to show (cf. section 4) that whenever

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v}) = \hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$$

then $\mathbf{v} = \mathbf{LinOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$ where $\mathbf{LinOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$ just returns the arithmetic mean of $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}$. Hence $\hat{\Delta}^{\mathrm{KL}}$ is a generalisation of the arithmetic pooling operator **LinOp**, and indeed coincides with that operator in the special case when each $\mathbf{K}_i$ specifies a unique probability function.

It is straightforward to prove that $\sum_{j \in \mathrm{Sig}(\mathbf{y})} x_j \log \frac{x_j}{y_j}$ is a convex function over the domain $\{(\mathbf{x}, \mathbf{y}) \in \mathbb{D}^L \times \mathbb{D}^L \colon \mathbf{y} \gg \mathbf{x}\}$. It follows that the set $\hat{\Delta}^{\mathrm{KL}}_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is nonempty, closed and convex for all $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ and hence that the p-merging operator $\hat{\Delta}^{\mathrm{KL}}$ satisfies **(K1)**. As in the case of Δ^{KL} , the fact that the operator $\hat{\Delta}^{\mathrm{KL}}$ satisfies **(K2)** follows at once from the remark above that $\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n} = 0$ if and only if $\mathbf{v} = \mathbf{w}^{(1)} = \ldots = \mathbf{w}^{(n)}$ in the definition of $\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$. Similarly $\hat{\Delta}^{\mathrm{KL}}$ satisfies **(K3)** trivially by definition.

It can also be shown that, as in the case of Δ^{KL}, Language Invariance and the Consistent Irrelevant Information Principle of section 3.1 also hold for the p-merging operator $\hat{\Delta}^{\mathrm{KL}}$. Finally if the "chairman" procedure of section 3.1, which related Δ^{KL} to **ME** is instead applied using $\hat{\Delta}^{\mathrm{KL}}$ then the point chosen in in $V_{\mathbf{K}}$ is not the maximum entropy point, but the $\mathbf{CM}^{\infty}$ point, or limit centre of mass point, of $V_{\mathbf{K}}$. These last results will be proved in a forthcoming paper.

Before stating our second theorem of this article we introduce the following natural strengthening of the disagreement principle **(K4)**.

(K4*) Strong Disagreement Principle.

Let $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ and $\mathbf{F}_1, \ldots, \mathbf{F}_m \in CL$.
Then $\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \Delta(\mathbf{F}_1, \ldots, \mathbf{F}_m) = \emptyset$ implies that

$$\Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n, \mathbf{F}_1, \ldots, \mathbf{F}_m) \cap \Delta(\mathbf{K}_1, \ldots, \mathbf{K}_n) = \emptyset.$$

Trivially the strong disagreement principle implies the disagreement principle.

Theorem 3.2. *The p-merging operator* $\hat{\Delta}^{\mathrm{KL}}$ *satisfies* **(K1)**, **(K2)**, **(K3)** *and* **(K5)**. *Furthermore the evidence bases to which* Δ^{KL} *is applied are restricted to BCL,* Δ^{KL} *satisfies the Strong Principle of Disagreement* **(K4*)**.

The fact that **(K1)**, **(K2)** and **(K3)** hold for $\hat{\Delta}^{\mathrm{KL}}$ has been established above. The proofs for **(K4*)** and **(K5)** will be given in the next section.

□

Historical Remarks.

Minimising Kullback-Leibler divergence from a convex set to a given probability function, or KL-projection, has long been used for updating and in machine learning algorithms (see e.g. [3], [6], [5], [8] and [2]). Connections between the minimisation of sums of Kullback-Leibler divergences and the operators **LinOp** and **LogOp** have also been noted previously by several authors within somewhat different frameworks. In particular we should mention the work of Matúš [13] who proved a number of convergence theorems covering the iteration of alternating operations of KL–projection or its dual to several compact convex sets followed by **LinOp** or, respectively, **LogOp**, and showing that under certain conditions these iterations converge to fixed points. These fixed points correspond respectively to particular points of $\hat{\Delta}^{\mathrm{KL}}$ or Δ^{KL}.

4 Proofs of Results

In this section we prove the two main results of this paper – the theorems 3.2 and 3.1. Since the properties **(K1)**, **(K2)** and **(K3)** have already been established for the two p-merging processes it remains to deal with the agreement and disagreement principles. The proofs of the agreement principle **(K5)** are straightforward and are given in 4.5 below. However the proofs for the disagreement principle (**(K4)** or **(K4*)**) are more complex and are different in flavour for Δ^{KL} and for $\hat{\Delta}^{\mathrm{KL}}$. The result for $\hat{\Delta}^{\mathrm{KL}}$ is proved in 4.6 and that for Δ^{KL} in 4.8.

We start by reviewing some geometrical properties of the space of probability functions $\mathbb{D}^L$ with respect to the divergence KL. First of all notice that for given $\mathbf{v} \in \mathbb{D}^L$ the Kullback-Leibler divergence $\mathrm{KL}(\mathbf{w}\|\mathbf{v})$ is a strictly convex function in the first argument over the domain specified by $\mathbf{v} \gg \mathbf{w}$. Owing to this if $\mathbf{v} \in \mathbb{D}^L$ is given and $W \subseteq \mathbb{D}^L$ is a closed convex set such that there is at least one probability function in W which $\mathbf{v}$ dominates, then we can define the KL-*projection* of $\mathbf{v}$ to W. This is defined as that unique point $\mathbf{w} \in W$ which minimizes $\mathrm{KL}(\mathbf{w}\|\mathbf{v})$. For more details see [2].

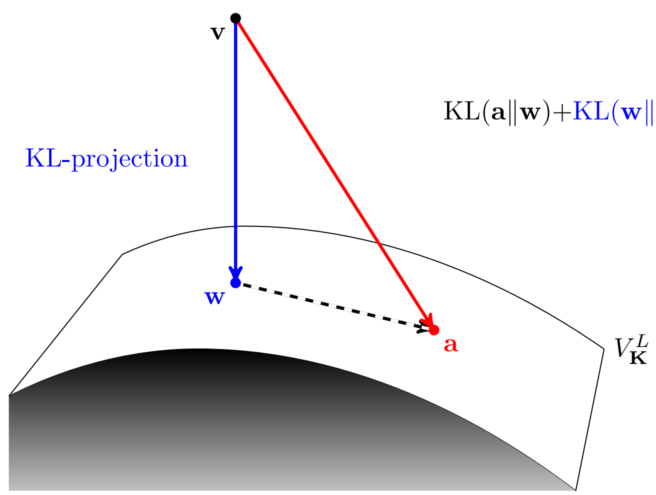
The following theorem is due to Csiszár [5].

Theorem 4.1 (Extended Pythagorean Theorem)**.** *Let* $\mathbf{w}$ *be the* KL-*projection of* $\mathbf{v} \in \mathbb{D}^L$ *to a closed convex set* $W \subseteq \mathbb{D}^L$*. Let* $\mathbf{a} \in W$ *be such that* $\mathbf{v} \gg \mathbf{w} \gg \mathbf{a}$*. Then*

$$\mathrm{KL}(\mathbf{a}\|\mathbf{w}) + \mathrm{KL}(\mathbf{w}\|\mathbf{v}) \leq \mathrm{KL}(\mathbf{a}\|\mathbf{v}).$$

□

An illustration of the extended Pythagorean theorem:



The following theorem is well known in information theory, see for instance [4].

Theorem 4.2 (Parallelogram Theorem)**.** *Let* $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}, \mathbf{v} \in \mathbb{D}^L$ *be such that* $\mathbf{v} \gg \mathbf{w}^{(i)}$ *for all* $1 \leq i \leq n$*. Then*

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)}\|\mathbf{v}) = \sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)}\| \mathbf{LinOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}))+$$

$$+n \cdot \mathrm{KL}(\mathbf{LinOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})\|\mathbf{v}).$$

□

Lemma 4.3. *Let* $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(n)} \in \mathbb{D}^L$ *and* $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(n)} \in \mathbb{D}^L$ *be such that*

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{u}^{(i)} \| \mathbf{v}) > \sum_{i=1}^{n} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{a}),$$

where $\mathbf{v} = \mathbf{LinOp}(\mathbf{u}^{(1)} \ldots, \mathbf{u}^{(n)})$ *and* $\mathbf{a} = \mathbf{LinOp}(\mathbf{a}^{(1)} \ldots, \mathbf{a}^{(n)})$. *Assume that* $\mathbf{u}^{(i)} \gg \mathbf{a}^{(i)}$ *for all* $1 \leq i \leq n$. *Then*

$$\sum_{i=1}^{n} \sum_{j \in \mathrm{Sig}(\mathbf{v})} (a_j^{(i)} - u_j^{(i)}) \cdot (\log u_j^{(i)} - \log v_j) < 0.$$

Proof. First of all notice that by the assumption $\mathbf{u}^{(i)} \gg \mathbf{a}^{(i)}$ for all $1 \leq i \leq n$ we have that

$$\mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{v}) - \mathrm{KL}(\mathbf{u}^{(i)} \| \mathbf{v}) - \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{u}^{(i)}) = \sum_{j \in \mathrm{Sig}(\mathbf{v})} (a_j^{(i)} - u_j^{(i)}) \cdot (\log u_j^{(i)} - \log v_j). \tag{3}$$

The above makes sense since $\mathbf{v} \gg \mathbf{u}^{(i)}$ for all $1 \leq i \leq n$. By the parallelogram theorem

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{v}) = \sum_{i=1}^{n} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{a}) + n \cdot \mathrm{KL}(\mathbf{a} \| \mathbf{v}).$$

Hence

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{a}) - \sum_{i=1}^{n} \mathrm{KL}(\mathbf{u}^{(i)} \| \mathbf{v}) + n \cdot \mathrm{KL}(\mathbf{a} \| \mathbf{v}) -$$

$$- \sum_{i=1}^{n} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{u}^{(i)}) = \sum_{i=1}^{n} \sum_{j \in \mathrm{Sig}(\mathbf{v})} (a_j^{(i)} - u_j^{(i)}) \cdot (\log u_j^{(i)} - \log v_j). \tag{4}$$

Since $\mathrm{KL}(\mathbf{w} \| \mathbf{v})$ is a convex function in both arguments whenever $\mathbf{v} \gg \mathbf{w}$, by the Jensen inequality

$$n \cdot \mathrm{KL}(\mathbf{a} \| \mathbf{v}) - \sum_{i=1}^{n} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{u}^{(i)}) \leq 0. \tag{5}$$

The inequality (5) together with the assumption that

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{u}^{(i)} \| \mathbf{v}) > \sum_{i=1}^{n} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{a})$$

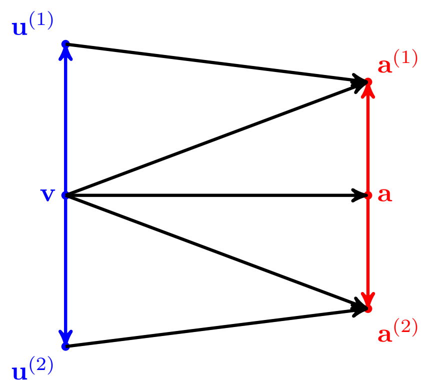
gives that left-hand side of the equality (4) is negative and so the right-hand side and

$$\sum_{i=1}^{n} \sum_{j \in \mathrm{Sig}(\mathbf{v})} (a_j^{(i)} - u_j^{(i)}) \cdot (\log u_j^{(i)} - \log v_j) < 0$$

follows.

□

The following picture illustrates the situation in the proof above for $n = 2$. Arrows indicate corresponding Kullback-Leibler divergences.



Lemma 4.4. *Let* $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)} \in \mathbb{D}^L$ *be fixed. Then*

(i) $\sum_{i=1}^n \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v})$ *is strictly minimal for* $\mathbf{v} = \mathbf{LinOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$.

(ii) $\sum_{i=1}^n \mathrm{KL}(\mathbf{v} \| \mathbf{w}^{(i)})$ *is strictly minimal for* $\mathbf{v} = \mathbf{LogOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$ *provided that* $\mathbf{LogOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$ *is defined, i.e. provided that for some* j *for all* i $\mathbf{w}_j^{(i)} \neq 0$.

Proof. *(i)* Let $\mathbf{w} = \mathbf{LinOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$. Clearly $\mathbf{w} \gg \mathbf{w}^{(i)}$ for all $1 \leq i \leq n$. It is easy to see that $\sum_{i=1}^n \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v}) - \sum_{i=1}^n \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{w}) = n \cdot \mathrm{KL}(\mathbf{w} \| \mathbf{v})$. Since $\mathrm{KL}(\mathbf{w} \| \mathbf{v}) = 0$ only if $\mathbf{v} = \mathbf{w}$ it follows that $\sum_{i=1}^n \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v})$ is strictly minimal for $\mathbf{v} = \mathbf{LinOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)})$.

(ii) The proof is equally straightforward, see e.g. [22].

$\square$

We will denote by $\hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ the set of all n-tuples $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}^L, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}^L$ such that for some $\mathbf{v} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$

$$\sum_{i=1}^n \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v}) = \hat{C}_{\mathbf{K}_1, \ldots, \mathbf{K}_n}.$$

This notation will be useful in the following two proofs.

Theorem 4.5. *(i) The* $\hat{\Delta}^{\mathrm{KL}}$ *p-merging operator satisfies* **(K5)**.

(ii) The Δ^{KL} *p-merging operator satisfies* **(K5)** *for all evidence bases in* $WBCL$.

Proof. The proofs are very similar in both cases, so we shall just give the proof for $\hat{\Delta}^{\mathrm{KL}}$ below.

Since we are assuming that $\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \ldots, \mathbf{F}_m) \neq \emptyset$, there is some $\mathbf{v} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \ldots, \mathbf{F}_m)$. For any such $\mathbf{v}$ this is equivalent to the assertion that for some $\mathbf{w}^{(1)} \ldots \mathbf{w}^{(n)} \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ and some $\mathbf{u}^{(1)} \ldots \mathbf{u}^{(m)} \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{F}_1, \ldots, \mathbf{F}_m)$

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v}) = \hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n} \ \text{ and } \ \sum_{i=1}^{m} \mathrm{KL}(\mathbf{u}^{(i)} \| \mathbf{v}) = \hat{C}_{\mathbf{F}_1,\ldots,\mathbf{F}_m}.$$

Then since by definition $\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n} + \hat{C}_{\mathbf{F}_1,\ldots,\mathbf{F}_m} \leq \hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n,\mathbf{F}_1,\ldots,\mathbf{F}_m}$ the same vectors $\mathbf{v}, \mathbf{w}^{(1)} \ldots \mathbf{w}^{(n)}, \mathbf{u}^{(1)} \ldots \mathbf{u}^{(m)}$ globally minimize the sum

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{w}^{(i)} \| \mathbf{v}) + \sum_{i=1}^{m} \mathrm{KL}(\mathbf{u}^{(i)} \| \mathbf{v}) \tag{6}$$

subject to $\mathbf{w}^{(i)} \in V_{\mathbf{K}_i}^L$, $1 \leq i \leq n$ and $\mathbf{u}^{(i)} \in V_{\mathbf{F}_i}^L$, $1 \leq i \leq m$.
Thus $\mathbf{v} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n, \mathbf{F}_1, \ldots, \mathbf{F}_m)$, and

$$\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n} + \hat{C}_{\mathbf{F}_1,\ldots,\mathbf{F}_m} \ = \ \hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n,\mathbf{F}_1,\ldots,\mathbf{F}_m}. \tag{7}$$

Since $\mathbf{v}$ was arbitrary we have proved that

$$\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \ldots, \mathbf{F}_m) \ \subseteq \ \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n, \mathbf{F}_1, \ldots, \mathbf{F}_m).$$

Now suppose $\mathbf{x} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n, \mathbf{F}_1, \ldots, \mathbf{F}_m)$. Then for some $\mathbf{y}^{(1)} \ldots \mathbf{y}^{(n)}$, $\mathbf{z}^{(1)} \ldots \mathbf{z}^{(m)} \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n, \mathbf{F}_1, \ldots, \mathbf{F}_m)$ and

$$\sum_{i=1}^{n} \mathrm{KL}(\mathbf{y}^{(i)} \| \mathbf{v}) + \sum_{i=1}^{m} \mathrm{KL}(\mathbf{z}^{(i)} \| \mathbf{v}) = \hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n,\mathbf{F}_1,\ldots,\mathbf{F}_m} \ .$$

In view of (7) if we did not now have that $\sum_{i=1}^{n} \mathrm{KL}(\mathbf{y}^{(i)} \| \mathbf{x}) = \hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$ and $\sum_{i=1}^{m} \mathrm{KL}(\mathbf{z}^{(i)} \| \mathbf{x}) = \hat{C}_{\mathbf{F}_1,\ldots,\mathbf{F}_m}$ then this would contradict the minimality of either $\hat{C}_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$ or $\hat{C}_{\mathbf{F}_1,\ldots,\mathbf{F}_m}$. Hence $\mathbf{x} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \ldots, \mathbf{K}_n) \cap \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \ldots, \mathbf{F}_m)$ and the result is proved.

The proof for Δ^{KL} is similar except that the final argument involving equation (7) fails if either of the quantities $C_{\mathbf{K}_1,\ldots,\mathbf{K}_n}$ or $C_{\mathbf{F}_1,\ldots,\mathbf{F}_m}$ is $+\infty$, which is the reason for the restriction of evidence bases to $WBCL$ in that case.

□

The following theorem proves that the $\hat{\Delta}^{\mathrm{KL}}$ p-merging operator satisfies the strong disagreement principle **(K4*)** if the evidence bases are restricted to BCL. This together with theorem 4.5 above and the results of section 3.2 is sufficient to establish our theorem 3.2.

Theorem 4.6. *Let* $\mathbf{K}_1, \dots, \mathbf{K}_n \in CL$ *and* $\mathbf{F}_1, \dots, \mathbf{F}_m \in CL$ *be such that for every*

$$(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}) \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m)$$

there is $(\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(m)}) \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m)$ *such that*

$$\mathbf{u}^{(i)} \gg \mathbf{a}^{(i)} \textit{ for all } 1 \le i \le m.$$

Then $\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n) \cap \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m) = \emptyset$ *implies*

$$\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m) \cap \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n) = \emptyset.$$

Proof. Assume that $\mathbf{v} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n)$ and

$$\mathbf{v} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m).$$

Let $(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)}) \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n)$ be an n-tuple associated with $\mathbf{v}$; in particular then $\mathbf{v} = \mathbf{LinOp}(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)})$.
Let

$$(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(n)}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}) \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m)$$

be an $(n+m)$-tuple associated with $\mathbf{v}$; then

$$\mathbf{v} = \mathbf{LinOp}(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(n)}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}).$$

This can only happen when

$$\mathbf{w}^{(i)} = \mathbf{v}^{(i)} \text{ for all } 1 \le i \le n$$

since the projections of the fixed $\mathbf{v}$ to each $V_{\mathbf{K}_i}$ are unique. Since in that case

$$\mathbf{v} = \mathbf{LinOp}(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)})$$

and

$$\mathbf{v} = \mathbf{LinOp}(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}).$$

we have that

$$\mathbf{v} = \mathbf{LinOp}(\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}).$$

Now let $\mathbf{a} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m)$ and $(\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(n)}) \in \hat{\Gamma}_L^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m)$ be associated m-tuple with $\mathbf{a}$ and such that $\mathbf{u}^{(i)} \gg \mathbf{a}^{(i)}$ for all $1 \le i \le m$. This is possible by the assumption of the theorem.

If $\mathbf{v} \in \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m)$ then

$$\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n) \cap \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m) \neq \emptyset$$

and we are done. On the other hand we show that $\mathbf{v} \notin \hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m)$ leads to a contradiction. First of all notice that from this assumption it follows that

$$\sum_{i=1}^{m} \mathrm{KL}(\mathbf{u}^{(i)} \| \mathbf{v}) > \sum_{i=1}^{m} \mathrm{KL}(\mathbf{a}^{(i)} \| \mathbf{a}).$$

Then by the lemma 4.3

$$\sum_{i=1}^{m}\sum_{j\in\mathrm{Sig}(\mathbf{v})}(a_j^{(i)}-u_j^{(i)})\cdot(\log u_j^{(i)}-\log v_j)<0.$$

On the other hand by the extended Pythagorean theorem (the theorem 4.1) and by the equation (3)

$$0\leq\sum_{i=1}^{m}\mathrm{KL}(\mathbf{a}^{(i)}\|\mathbf{v})-\mathrm{KL}(\mathbf{u}^{(i)}\|\mathbf{v})-\mathrm{KL}(\mathbf{a}^{(i)}\|\mathbf{u}^{(i)})=$$

$$=\sum_{i=1}^{m}\sum_{j\in\mathrm{Sig}(\mathbf{v})}(a_j^{(i)}-u_j^{(i)})\cdot(\log u_j^{(i)}-\log v_j)<0.$$

which is a contradiction.

□

The following counterexample shows that the assumption of theorem 4.6 restricting evidence bases to BCL is necessary even if we reformulate the theorem using the weaker disagreement principle **(K4)** in place of **(K4*)**.

Example 4.7. Assume that $|L|=2$, $V_{\mathbf{K}_1}=\{(1,0,0,0)\}$, $V_{\mathbf{K}_2}=\{(0,1,0,0)\}$, $V_{\mathbf{F}_1}=\{(x,0,1-x,0)\,:\,x\in[0,1]\}$ and $V_{\mathbf{F}_2}=\{(0,x,1-x,0)\,:\,x\in[0,1]\}$. Clearly $\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1,\mathbf{K}_2)=\{(\frac{1}{2},\frac{1}{2},0,0)\}$ and $\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1,\mathbf{F}_2)=\{(0,0,1,0)\}$. Therefore $\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1,\mathbf{K}_2)\cap\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{F}_1,\mathbf{F}_2)=\emptyset$. It can now be shown that

$$\hat{\Delta}_L^{\mathrm{KL}}(\mathbf{K}_1,\mathbf{K}_2,\mathbf{F}_1,\mathbf{F}_2)=\{(\tfrac{1}{2},\tfrac{1}{2},0,0)\}$$

which suffices to contradict the disagreement principle.

□

Before leaving the discussion of $\hat{\Delta}^{\mathrm{KL}}$ we note that this p-merging operator is one of a large class of p-merging operators which all satisfy the same properties as $\hat{\Delta}^{\mathrm{KL}}$ does in Theorem 4.6. These are formed by the class of operators generated by substituting any convex Bregman divergence[12] in place of Kullback–Leibler divergence in the definition of $\hat{\Delta}^{\mathrm{KL}}$. This holds primarily because the well-known geometric properties of Bregman divergences, such as the extended Pythagorean theorem above, are exactly what is required for the proof of **(K4*)**. Amongst such Bregman divergences is the very special case of squared Euclidean distance E2 defined by

$$\mathrm{E2}(\mathbf{w}\|\mathbf{v})=\sum_{j=1}^{J}(w_j-v_j)^2.$$

[12]For the definition of a Bregman divergence see [2].

Since in the case of this divergence the zero points cause no discontinuity, the strong disagreement principle holds without any restriction on the class CL for the $\hat{\Delta}^{\mathrm{E2}}$ p-merging operator defined for any $\mathbf{K}_1, \dots, \mathbf{K}_n \in CL$ as the set $\hat{\Delta}^{\mathrm{E2}}_L(\mathbf{K}_1, \dots, \mathbf{K}_n)$ of probability functions $\mathbf{v} \in \mathbb{D}^L$ which globally minimise the sum of squared Euclidean distances

$$\sum_{i=1}^{n} \mathrm{E2}(\mathbf{w}^{(i)} \| \mathbf{v}) \tag{8}$$

subject only to the conditions that $\mathbf{w}^{(1)} \in V^L_{\mathbf{K}_1}, \dots, \mathbf{w}^{(n)} \in V^L_{\mathbf{K}_n}$.

However what all the p-merging operators defined by convex Bregman divergences have in common is that they are generalisations of **LinOp** and reduce to **LinOp** when marginalised as pooling operators. The social entropy operator Δ^{KL}, which marginalises to the **LogOp** pooling operator therefore has very different characteristics from p-merging operators defined in this way.

Finally the theorem below proves that the Δ^{KL}-merging operator satisfies the disagreement principle **(K4)** if the evidence bases are restricted to $WBCL$. This together with theorem 4.5 above, and the earlier results of section 3.1, is sufficient to establish our theorem 3.1.

Theorem 4.8. *For all evidence bases* $\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m \in WBCL$ *the social entropy operator* Δ^{KL} *satisfies* (**K4**).

Proof. Let $\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m \in WBCL$ be such that $\bigcap_{i=1}^m V_{\mathbf{F}_i} \neq \emptyset$. Given our assumption that evidence bases belong to $WBCL$ we can assume that for every $1 \leq i \leq n$ and every $1 \leq j \leq J$ there is $\mathbf{w} \in V_{\mathbf{K}_i}$ such that $w_j \neq 0$. Then we must show that

$$\Delta^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n) \cap \Delta^{\mathrm{KL}}(\mathbf{F}_1, \dots, \mathbf{F}_m) = \emptyset$$

implies that

$$\Delta^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m) \cap \Delta^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n) = \emptyset.$$

We prove the contrapositive. Suppose that for some fixed $\mathbf{v}$ we have that $\mathbf{v} \in \Delta^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n)$ and $\mathbf{v} \in \Delta^{\mathrm{KL}}(\mathbf{K}_1, \dots, \mathbf{K}_n, \mathbf{F}_1, \dots, \mathbf{F}_m)$. Let $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)}$ be such that they minimize $\sum_{j=1}^J \sum_{i=1}^n v_j \log \frac{v_j}{v_j^{(i)}}$ subject to $\mathbf{v}^{(1)} \in V_{\mathbf{K}_1}, \dots, \mathbf{v}^{(n)} \in V_{\mathbf{K}_n}$. Then

$$\mathbf{v} = \mathbf{LogOp}(\mathbf{v}^{(1)} \dots, \mathbf{v}^{(n)}). \tag{9}$$

Similarly let $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(n)}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}$ be such that they minimize

$$\sum_{j=1}^{J} \sum_{i=1}^{n} v_j \log \frac{v_j}{w_j^{(i)}} + \sum_{j=1}^{J} \sum_{i=1}^{m} v_j \log \frac{v_j}{u_j^{(i)}}$$

subject to

$$\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n} \text{ and } \mathbf{u}^{(1)} \in V_{\mathbf{F}_1}, \ldots, \mathbf{u}^{(m)} \in V_{\mathbf{F}_m}.$$

Equivalently $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}, \mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}$ are such as to maximize

$$\sum_{j=1}^{J} \Big[\prod_{i=1}^{n} w_j^{(i)} \prod_{i=1}^{m} u_j^{(i)} \Big]^{\frac{1}{n+m}} \tag{10}$$

subject to $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$ and $\mathbf{u}^{(1)} \in V_{\mathbf{F}_1}, \ldots, \mathbf{u}^{(m)} \in V_{\mathbf{F}_m}$ and are such that

$$\mathbf{v} = \mathbf{LogOp}(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}, \mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}). \tag{11}$$

The above can only happen when for all j such that $v_j \neq 0$

$$w_j^{(i)} = v_j^{(i)} \text{ for all } 1 \leq i \leq n \tag{12}$$

since $\sum_{j, v_j \neq 0} v_j \log \frac{v_j}{w_j^{(i)}}$ for fixed $\mathbf{v}$ is strictly convex function and hence it has a unique minimizer subject to $\mathbf{w}^{(i)} \in V_{\mathbf{K}_i}$ for all $1 \leq i \leq n$. Notice that this holds only under the assumption that for every $1 \leq i \leq n$ there is $\mathbf{x} \in V_{\mathbf{K}_i}$ such that $x_j \neq 0$. Hence by (9) and (10) the equation (11) can be rewritten as

$$v_j = \frac{[\prod_{i=1}^{m} u_j^{(i)}]^{\frac{1}{m}}}{\frac{(\sum_j [\prod_{i=1}^{n} v_j^{(i)} \prod_{i=1}^{m} u_j^{(i)}]^{\frac{1}{n+m}})^{\frac{m+n}{m}}}{(\sum_j [\prod_{i=1}^{n} v_j^{(i)}]^{\frac{1}{n}})^{\frac{n}{m}}}}.$$

Notice that the denominator is independent of j.

On the other hand if for any given j $v_j = 0$ then $w_j^{(1)} = 0, \ldots, w_j^{(n)} = 0$ and $u_j^{(1)} = 0, \ldots, u_j^{(m)} = 0$. This is proved in [23] theorem 3.6 (ii), and holds only under the assumption that for every $1 \leq i \leq n$ and every j there is $\mathbf{x} \in V_{\mathbf{K}_i}$ such that $x_j \neq 0$. So the denominator above is certainly non-zero.

Putting the above together it follows that

$$\mathbf{v} = \mathbf{LogOp}(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}).$$

Now let $\mathbf{a}$ be consistent with $\mathbf{F}_i$ for all i so that in particular

$$\mathbf{a} \in \Delta^{\mathrm{KL}}(\mathbf{F}_1, \ldots, \mathbf{F}_m) = \bigcap_{i=1}^{m} V_{\mathbf{F}_i}.$$

Consider

$$F(\lambda) = \sum_{j=1}^{J} \Big[\prod_{i=1}^{n} v_j^{(i)} \prod_{i=1}^{m} (u_j^{(i)} + \lambda(a_j - u_j^{(i)})) \Big]^{\frac{1}{n+m}}.$$

We need to show that

$$\Delta^{\mathrm{KL}}(\mathbf{K}_1,\dots,\mathbf{K}_n)\cap\Delta^{\mathrm{KL}}(\mathbf{F}_1,\dots,\mathbf{F}_m)=\Delta^{\mathrm{KL}}(\mathbf{K}_1,\dots,\mathbf{K}_n)\cap\bigcap_{i=1}^{m}V_{\mathbf{F}_i}\neq\emptyset.$$

For this it is sufficient to prove that $\frac{d}{d\lambda}F|_{\lambda=0}>0$ unless $\mathbf{u}^{(1)}=\dots=\mathbf{u}^{(n)}=\mathbf{v}$. The idea here is that if the maximum value of F is obtained for $\mathbf{u}^{(1)},\dots,\mathbf{u}^{(n)}$ which are not all equal then from the existence of the point $\mathbf{a}\in\bigcap_{i=1}^{m}V_{\mathbf{F}_i}$ we can show that $F(0)<F(\lambda)$ for some $\lambda>0$; this is a contradiction since by the convexity of the $V_{\mathbf{F}_i}$ each $\mathbf{u}^{(i)}+\lambda(\mathbf{a}-\mathbf{u}^{(i)})\in V_{\mathbf{F}_i}$ so that $F(0)<F(\lambda)$ contradicts the maximality of (10) given (12).

Note that $v_j=\frac{(\prod_{i=1}^{m}u_j^{(i)})^{\frac{1}{m}}}{M}$ where $M=\sum_{j=1}^{J}(\prod_{i=1}^{m}u_j^{(i)})^{\frac{1}{m}}<1$ unless $\mathbf{u}^{(1)}=\dots=\mathbf{u}^{(m)}$.

$$\frac{d}{d\lambda}F(\lambda)=\sum_{j=1}^{J}\Big(\prod_{i=1}^{n}v_j^{(i)}\Big)^{\frac{1}{n+m}}\Big[\prod_{i=1}^{m}u_j^{(i)}+\lambda\sum_{i=1}^{m}(a_j-u_j^{(i)})\prod_{i\neq i',i'=1,\dots,m}u_j^{(i')}+$$

$$+O(\lambda^2)\Big]^{\frac{1}{n+m}-1}\cdot\Big[\sum_{i=1}^{m}(a_j-u_j^{(i)})\prod_{i\neq i',i'=1,\dots,m}u_j^{(i')}+O(\lambda)\Big].$$

We got this by expanding $\prod_{i=1}^{m}(u_j^{(i)}+\lambda(a_j-u_j^{(i)}))=\prod_{i=1}^{m}u_j^{(i)}+\lambda\sum_{i=1}^{m}(a_j-u_j^{(i)})\prod_{i\neq i',i'=1,\dots,m}u_j^{(i')}+O(\lambda^2)$. Furthermore

$$\frac{d}{d\lambda}F(\lambda)|_{\lambda=0}=\sum_{j=1}^{J}(\prod_{i=1}^{n}v_j^{(i)})^{\frac{1}{n+m}}[(Mv_j)^m]^{-1+\frac{1}{n+m}}\cdot$$

$$\cdot\Big[\sum_{i=1}^{m}\Big(\frac{a_j}{u_j^{(i)}}M^m(v_j)^m-M^m(v_j)^m\Big)\Big]=C\cdot\sum_{j=1}^{J}(v_j)^{\frac{n}{n+m}-m+\frac{m}{n+m}+m}\Big[\sum_{i=1}^{m}\frac{a_j}{u_j^{(i)}}-1\Big]=$$

$$=C\cdot\Big[\sum_{j=1}^{J}\sum_{i=1}^{m}\frac{v_ja_j}{u_j^{(i)}}-m\Big]=C\cdot\Big[\sum_{j=1}^{J}a_j\sum_{i=1}^{m}\frac{(\prod_{k=1}^{m}u_j^{(k)})^{\frac{1}{m}}}{u_j^{(i)}M}-m\Big],$$

where $C=M^{\frac{1}{n+m}}\cdot(\sum_{j=1}^{J}[\prod_{i=1}^{n}v_j^{(i)}]^{\frac{1}{n}})^{\frac{n}{n+m}}$ is a positive constant.

Note that if $u_j^{(i)}=0$ and $a_j\neq 0$ for some $1\leq i\leq m$ and some j then $F(\lambda)\to+\infty$ as $\lambda\to 0^+$. On the other hand if also $a_j=0$ then we can just leave out that index j from the summation. Finally by the AG-inequality

$$C\cdot\Big[\sum_{j=1}^{J}a_j\sum_{i=1}^{m}\frac{1}{u_j^{(i)}}\frac{(\prod_{k=1}^{m}u_j^{(k)})^{\frac{1}{m}}}{M}-m\Big]\geq$$

$$\geq C \cdot \Big[\sum_{j=1}^{J} a_j m \cdot \Big(\prod_{i=1}^{m} \frac{1}{u_j^{(i)}}\Big)^{\frac{1}{m}} \frac{(\prod_{k=1}^{m} u_j^{(k)})^{\frac{1}{m}}}{M} - m\Big] = Cm\frac{1-M}{M}.$$

The last term is greater than 0 unless $\mathbf{u}^{(1)} = \ldots = \mathbf{u}^{(m)}$ which concludes the proof.

□

The following counterexample shows that the theorem above fails if $WBCL$ is replaced by CL in the theorem above.

Example 4.9. Let $V^L_{\mathbf{K}_1} = \{(0, 0, \frac{1}{3}, \frac{2}{3})\}$ and $V^L_{\mathbf{F}_1} = \{(0, \frac{1}{3}, \frac{2}{9}, \frac{4}{9})\}$.
Obviously $\Delta^{\mathrm{KL}}(\mathbf{K}_1) \cap \Delta^{\mathrm{KL}}(\mathbf{F}_1) = \emptyset$.
However $\Delta^{\mathrm{KL}}(\mathbf{K}_1, \mathbf{F}_1) = \mathbf{LogOp}\Big[(0, 0, \frac{1}{3}, \frac{2}{3})\,,\,(0, \frac{1}{3}, \frac{2}{9}, \frac{4}{9})\Big] \;=\; (0, 0, \frac{1}{3}, \frac{2}{3})$.

□

References

[1] Adamčík, M. and Wilmers, G.M. (2012), *The Irrelevant Information Principle for Collective Probabilistic Reasoning*, Proceedings of the 9th Workshop on Uncertain Processing, Mariánské Lázně, pp. 1-12

[2] Amari, S. (2009), *Divergence, Optimization and Geometry*, Neural Information Processing: 16th International Conference, Iconip

[3] Collins, M. and Schapire, R.E. (2002), *Logistic Regression, AdaBoost and Bregman Distances*, Machine Learning, 48, pp. 253-285

[4] Cover, T.M. and Thomas, J.A. (1991), *Elements of Information Theory*, Wiley Series in Telecommunications, John Wiley and Sons, New York

[5] Csiszár, I. (1975), *I-Divergence Geometry of Probability Distribution and Minimization Problems*, The Annals of Probability, Vol. 3, No. 1, pp. 146-158

[6] Csiszár, I. and Tusnády, G. (1984), *Informational Geometry and Alternating Minimization Procedures*, Statistic and Decisions, Supplement Issue No. 1, pp. 205-237

[7] Deming, W.E. and Stephan, F.F. (1940), *On a least square adjustment of a sampled frequency table when the expected marginals totals are unknown*, Ann. Math. Statist., 11:427-444

[8] Hájek, P., Havránek, T. and Jiroušek, J. (1992), *Uncertain Information Processing in Expert Systems*, CRC Press, Boca Raton, Ann Arbor, London, Tokyo

[9] Genest, C. and Zidek, J.V. (1986), *Combining probability distributions: A critique and an annotated bibliography.* Statistical Science Vol 1(1) pp. 114-135.

[10] Kern-Isberner, G. and Rödder, W. (2004), *Belief Revision and Information Fusion on Optimum Entropy*, International Journal of Intelligent Systems, 19, pp 837-857,

[11] Konieczny, S. and Pino Pérez, R. (1998) *On the Logic of Merging*, Proceedings of the 6th International Conference on Principles of Knowledge Representation and Reasoning, pp. 488-98, Morgan Kaufman, San Francisco, CA

[12] Konieczny, S. and Pino Pérez, R. (2011) *Logic Based Merging*, J. Philosophical Logic 40, pp. 239-270.

[13] Matúš, F. (2007), *On iterated averages of I-projections*, Universitat Bielefeld, Germany

[14] Osherson, D. and Vardi, M. (2006), *Aggregating disparate estimates of chance*, Games and Economic Behavior 56(1) pp. 148-173, July 2006.

[15] Paris, J.B. (1994), *The uncertain reasoner companion*, Cambridge University Press, Cambridge

[16] Paris, J.B. and Vencovská, A. (1989), *On the Applicability of Maximum Entropy to Inexact Reasoning*, International Journal of Approximate Reasoning, 3

[17] Paris, J.B. and Vencovská, A. (1990), *A Note on the Inevitability of Maximum Entropy*, International Journal of Approximate Reasoning, 4, pp. 183-224

[18] Predd, J.B., Osherson, D.N., Kulkarni, S.R., and Vincent Poor, H. (2008), *Aggregating Probabilistic Forecasts from Incoherent and Abstaining Experts*, Decision Analysis, 5(4) pp. 177-189

[19] Williamson, J. (2009), *Aggregating Judgements by Merging Evidence* ,J Logic Computation 19(3): 461-473

[20] Williamson, J. (2010), *In Defense of Objective Bayesianism*, Oxford University Press

[21] Williamson, J. (2013), *Deliberation, Judgement and the Nature of Evidence*, unpublished manuscript.

[22] Wilmers, G.M. (2010), *The Social Entropy Process: Axiomatising the Aggregation of Probabilistic Beliefs* , Probability, Uncertainty and Rationality, pp 87-104, edited by Hykel Hosni and Franco Montagna, 10 CRM series, Scuola Normale Superiore, Pisa

[23] Wilmers, G.M. (2011), *Generalising the Maximum Entropy Inference Process to the Aggregation of Probabilistic Beliefs*, available from http://manchester.academia.edu/GeorgeWilmers/Papers