Investigation of Properties of Some Inference Processes

Hawes, Peter

2007

MIMS EPrint: **2009.53**

ISSN 1749-9097

INVESTIGATION OF PROPERTIES OF SOME INFERENCE PROCESSES

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2007

Peter Hawes
School of Mathematics

Contents

Word count 32767

The University of Manchester

Peter Hawes
Doctor of Philosophy
Investigation of properties of some inference processes
November 6, 2007

The spectrum of Renyi inference processes in the discrete case is found to have limits of $Minimax$ at one end and CM_∞ at the other. Another sequence of processes is found to have the limit $Maximin$. Although $Maximin$ is the dual of $Minimax$, it is seen to have better characteristics when compared with $Maximum\ Entropy$ (ME) than those possessed by $Minimax$. The comparison of inference processes is made using a list of desiderata which were shown by Paris/Vencovska to uniquely characterise ME.

Algorithms are described for calculating $Minimax$ and $Maximin$, which have the advantage over ME of inferring belief values which are rational numbers when the agent's knowledge is itself expressed purely in terms of rational numbers.

Then $Minimax$ and $Maximin$ are viewed as examples of $Partly\ Linear$, or PL inference processes. This yields a unique characterisation of $Maximin$. Another inference process, $Meanimax$, compares well with $Minimax$ and is a counterexample of some plausible conjectures about certain properties of inference processes.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

Thanks are due to my parents for their continuous support. The EPSRC have funded my work with a DTA scholarship. Also Prof. Jeff Paris introduced me to logic and Dr. George Wilmers, my supervisor, gave me much helpful advice. Other people have contributed also, in varying degrees, towards making my period of study less painful.

The Author

The author has a Masters of Mathematics degree from the University of Manchester. His research experience consists entirely of the work done for this thesis.

Chapter 1

Introduction

1.1 General notation

In this section we introduce notation that will be used generally in this thesis. In this thesis we investigate properties of inference processes on finite languages of propositional logic. A finite *language* L is a finite set of *propositional variables*, say $p_1, p_2, \ldots p_n$. The list of logical connectives we use to make *sentences* is

$$\rightarrow, \wedge, \vee, \neg \tag{1.1}$$

The set of sentences of the language L is denoted by SL.

We define an *atom* of L to be a sentence of the form

$$\alpha = p_1^{\epsilon_1} \wedge (p_2^{\epsilon_2} \wedge \ldots p_n^{\epsilon_n}))) \ldots) \tag{1.2}$$

where $\epsilon_i = 0$ or 1 for each $i = 1, 2, \ldots n$ and p_i^1 denotes p_i while p_i^0 denotes $\neg p_i$. Thus each atom α specifies the truth or falsity of every p.v. (propositional variable) and there are 2^n atoms. When the context only demands that sentences are defined up to logical equivalence, which is almost always the case, we write the atom above more informally as

$$\alpha = \bigwedge_{i=1}^{n} p_i^{\epsilon_i} \tag{1.3}$$

The set of atoms of L is denoted by $At^L = \{\alpha_i \,|1 \leq i \leq 2^n\}$, and, unless we state otherwise, J denotes $|At^L|$.

It will often be useful to have a standard enumeration of At^L. In the *standard ordering* of the atoms,

$$\alpha_{1+(\sum_{i=1}^{n} 2^{n-i}(1-\epsilon_i))} = \bigwedge_{i=1}^{n} p_i^{\epsilon_i} \tag{1.4}$$

so that, for example, $\alpha_1 = \bigwedge_{i=1}^{n} p_i$ and $\alpha_{2^n} = \alpha_J = \bigwedge_{i=1}^{n} \neg p_i$.

We let log denote the function that takes logarithms to base 2. Later in this introduction, and in Chapter 3, we use $\log_e$ to denote the natural logarithm function. We use $\mathbb{N}$ to denote the set of natural numbers, where a natural number is defined to be a *strictly* positive integer, so $0 \notin \mathbb{N}$. We use $\mathbb{Q}$ and $\mathbb{R}$ to denote the sets of rational numbers and real numbers, respectively.

Throughout this work the extent to which a rational agent believes that sentences $\theta \in SL$ are true are thought of as probabilities. We sometimes say that atoms of L have certain belief values etc. We may call these values the agent's beliefs, beliefs inferred, or our beliefs—the reader is credited with being a rational individual!

Belief as probability is controversial and we summarise some arguments in support of this philosophy at the beginning of the next section. The beliefs are given by *probabilistic belief functions*, called p.b. functions

$$Bel : SL \longrightarrow [0, 1] \tag{1.5}$$

which by definition, as in [Par], must satisfy the axioms:

(P1) If $\models \theta$ then $Bel(\theta) = 1$.

(P2) If $\models \neg(\theta \wedge \phi)$ then $Bel(\theta \vee \phi) = Bel(\theta) + Bel(\phi)$.

We can now deduce, by Disjunctive Normal Form, that each Bel on the language L is uniquely determined by its values on the atoms, $Bel(\alpha_i)$, which we usually denote by x_i [1] . We might call x_i the belief (value) of, or for α_i, or the belief in α_i, or the belief given to α_i by Bel. Where $J = 2^n$, the set of p.b. functions on L is thus

[1] This usage of Disjunctive Normal Form to express p.b. functions in terms of their values at the atoms of L is standard notation for work on inference processes—see [Moh], [Par] etc. However, there exist other parameterisations for p.b. functions. In information geometry the logarithms of the beliefs of the atoms are commonly used. In work on probabilistic networks it is common to use, for each $j = 1, \ldots n$, the conditional probabilities of the j'th propositional (continued on next page)

identified with

$$\mathbb{D}^J = \left\{ \vec{x} \in \mathbb{R}^J \mid \sum_{i=1}^{J} x_i = 1 \ \text{ and every } \ x_i \geq 0 \right\} \tag{1.6}$$

A *knowledge base* K consists of a finite set of linear equations, called *constraints*, which *may* be written in the form

$$\left\{ \sum_{j=1}^{J} a_{pj} Bel(\theta_j) = b_p \,|\, 1 \leq p \leq m \right\} \tag{1.7}$$

including the constraint $\sum_{i=1}^{J} Bel(\alpha_i) = 1$. However, where x_i denotes $Bel(\alpha_i)$ we can also express a knowledge base as a finite set of equations of the form

$$\left\{ \sum_{j=1}^{J} c_{pj} x_i = d_p \,|\, 1 \leq p \leq m \right\} \tag{1.8}$$

including $\sum_{i=1}^{J} x_i = 1$, where the language L and labelling of the atoms as $\alpha_1, \ldots \alpha_J$ are understood. Any knowledge base can be equivalently expressed in either form, using the identity

$$Bel(\theta) = \sum_{\alpha_i \models \theta} Bel(\alpha_i) \tag{1.9}$$

We usually only consider *consistent* knowledge bases, i.e. those satisfied by some p.b. function. We let CL denote the set of consistent knowledge bases over a language L. The solution set of K is denoted by $V^L(K)$ and is the set of those non-negative $\vec{x} \in \mathbb{R}^J$ satisfying the constraints of K; this is a set of p.b. functions viewed as vectors in $\mathbb{R}^J$, indeed in $\mathbb{D}^J$.

If K is a knowledge base we always make the following assumption, due to [Par], known as the *Watts Assumption*:

$$K \ \text{ is all the agent's relevant knowledge} \tag{1.10}$$

We also use the more general notion of a *constraint set*. A constraint set K on J co-ordinates is a set of linear constraints of the form of (1.8) on variables $x_1, x_2, \ldots x_J$, where J is an integer at least 2 and for which the set of solutions, denoted by $V(K)$,

variable, given that all of the truth values of the first $j-1$ p.v.'s are fixed. If different parameterisations were used, we could define inference processes with similar definitions to $Minimax^L$, $Maximin^L$ (see Chapters 3,5 resp.) etc. but they would not be the same inference processes.

which includes just those non-negative $\vec{x} \in \mathbb{R}^J$ satisfying the constraints of K, is bounded. There is not necessarily an association of the x_is as beliefs of atoms of a language; indeed J might not even be a power of 2.

When our knowledge is given by some $K \in CL$, we choose a p.b. function satisfying K by using an *inference process* N^L which is a function

$$N^L : CL \to \mathbb{D}^J \tag{1.11}$$

such that for all consistent K, $N^L(K)$ satisfies K, i.e. $N^L(K) \in V^L(K)$.

The set of *generalised* solutions of a constraint set or knowledge base K is denoted by $G(K)$. It is given by

$$G(K) = \{\vec{x} \in \mathbb{R}^J |\, \vec{x} \text{ satisfies the constraints of } K\} \tag{1.12}$$

so that $V(K) = \{\vec{x} \in G(K) \text{ s.t. } x_i \geq 0 \text{ for each } i = 1, \ldots J\}$. Also it is useful to define

$$I(K) = \{i \in \{1, 2, \ldots J\} \text{ s.t. } x_i = 0 \text{ for all } \vec{x} \in V(K)\} \tag{1.13}$$

for any constraint set K. We say a co-ordinate i is K-*constant* (abbreviated to "constant" when K is clear from the context) or *constant w.r.t.* K iff there exists $C \in \mathbb{R}$ such that $x_i = C$ for all $\vec{x} \in V(K)$.

Note that, if K is a knowledge base in CL, we may sometimes write $G(K)$ and $I(K)$ as $G^L(K)$ and $I^L(K)$ respectively, to make L explicit.

The following definitions, together with an equivalence result, give us characterisations for solution sets $V(K)$:

We say that a subset S of $\mathbb{R}^J$ is *line-bounded* iff no infinite line is a subset of S. A *convex facet-polytope* is a *line-bounded* set $P \subseteq \mathbb{R}^J$ expressible in the form:

$$P = \bigcap_{i=1}^{m} \left\{ \vec{x} \in \mathbb{R}^J \mid \sum_{j=1}^{J} a_{ij} x_j \leq b_i \right\} \tag{1.14}$$

where the a_{ij}'s and b_i's are real constants.

A *convex polytope* is a set $Q \subseteq \mathbb{R}^J$ that can be expressed in the form:

$$Q = \left\{ \sum_{i=1}^{m} a_i \vec{u}^{(i)} \, | a_i \geq 0 \text{ for each } i = 1, \dots m \text{ and } \sum_{i=1}^{m} a_i = 1 \right\}. \tag{1.15}$$

where the $\vec{u}^{(i)}$'s are constants. We say that Q is the set of *convex combinations* of the $\vec{u}^{(i)}$'s. In an expression of the form (1.15), a particular $\vec{u}^{(i*)}$ is *redundant* iff Q also equals the set of convex combinations of the $\vec{u}^{(i)}$'s apart from $\vec{u}^{(i*)}$. We see that $\vec{u}^{(i*)}$ is redundant iff it can be expressed as a convex combination of the other $\vec{u}^{(i)}$'s. If, in (1.15), none of the $\vec{u}^{(i)}$'s are redundant, the $\vec{u}^{(i)}$'s are known as the *vertices of Q w.r.t. (1.15).*

The following useful technical lemma can be found in, for example, [Zie]page 29:

Lemma 1 *The set of vertices of a convex polytope P w.r.t. an expression of the form* (1.15) *is independent of the choice of equation; they are now known as the* vertices of P. *For all $J \in \mathbb{N}$, the collection of all convex polytopes in $\mathbb{R}^J$ is the same as that of all convex facet-polytopes in $\mathbb{R}^J$.*

Remark By definition of a constraint set, $V(K)$ is a convex facet-polytope for every constraint set K. By Lemma 1, $V(K)$ is also a convex polytope.

Now, by definition of vertices, every vertex of $V(K)$ is on the topological boundary of $V(K)$ w.r.t. the subspace topology of $V(K)$. Hence, if K is given in the form

$$K = \left\{ \sum_{j=1}^{J} c_{pj} x_i = d_p \text{ for each } p = 1, \dots m \right\} \tag{1.16}$$

such that $G(K)$ does not merely consist of a single point [2], then every vertex of $V(K)$ is a solution of a constraint set of the form $K' = K \cup \{x_i = 0\}$, for some $i = 1, \dots$ or J.

The two lemmas which follow are known results about convex polytopes (see [Zie]) and we prove them here using the notation of this thesis.

[2] If $G(K) = \{\vec{X}\}$, say, either K is inconsistent or $\vec{X}$ is non-negative and is the only vertex of $V(K)$.

Lemma 2 *Let P be a convex polytope such that*

$$P = \bigcap_{i=1}^{m} \left\{ \vec{x} \in \mathbb{R}^J \mid \sum_{j=1}^{J} a_{ij} x_j \leq b_i \right\} \tag{1.17}$$

where the a_{ij}'s and b_i's are real constants. Let $C = \{\vec{x} \in \mathbb{R}^J \quad s.t. \quad \sum_{j=1}^{J} a_{1j} x_j = b_1\}$ and let $Q = P \cap C$.

Then the vertices of Q are precisely those vertices of P lying in Q.

Proof We assume w.l.o.g. that $Q \neq P$. When we express a point in Q as a convex combination of vertices of P, those vertices, $\vec{u}$, of P which we use must lie in Q since the value of $\sum_{j=1}^{J} a_{1j} u_j$ is less than b_1 at all of the other vertices of P. Thus every point in Q is a convex combination of the members of $Vert$, which denotes the set of vertices of P which lie in Q.

Also, if some of those vertices were redundant for describing Q, they would be expressible as a convex combination of other members of $Vert$. In that case, they would be redundant for describing P. Hence there are no such redundancies and we have proved the lemma.

$\blacksquare$

Lemma 3 *Let P be a convex polytope as in Lemma 2. The vertices of P are all unique solutions of sets of equations of the form*

$$P + \left\{ \sum_{j=1}^{J} a_{i_k j} x_j = b_{i_k} \;\; s.t. \;\; k = 1, \ldots d \right\} \tag{1.18}$$

where $i_1, i_2 \ldots i_d \in \{1, \ldots m\}$.

Proof The vertices of P lie on the topological boundary of P, using the subspace topology of the smallest affine set that contains P. Hence every vertex lies in a set that we can describe by changing one of the inequalities of (1.17) into an equality. Using Lemma 2, the proof follows by induction on the dimension of the smallest affine space that contains P.

$\blacksquare$

The following lemma is a useful existence result which we use later in this thesis.

Lemma 4 *Suppose that L is an overlying language and that $V \subseteq \mathbb{D}^J$ is the intersection of an affine subset of $\mathbb{R}^J$ with $\mathbb{D}^J$. Then there exists a knowledge base K such that $V = V^L(K)$.*

Proof Let $V = \mathbb{D}^J \cap G$, where G is affine. Now we write a basis of the vector space of those $\vec{x} \in \mathbb{R}^J$ such that $\vec{x} \cdot (\vec{g}_1 - \vec{g}_2) = 0$ for all $\vec{g}_1, \vec{g}_2 \in G$—the vectors perpendicular to G. Each member of this basis gives rise to a linear equation in the x_i's and when these equations are considered together, G is precisely the set of vectors in $\mathbb{R}^J$ that satisfy all of those equations. We include the equation $\sum_{i=1}^J x_i = 1$ and call the equations constraints. We have now written K such that $V^L(K) = V$, proving the lemma.

∎

Definition Let S be some convex subset of $\mathbb{R}^J$, where J is an integer. Suppose that a function $f : S \to \mathbb{R}$ is such that for all distinct vectors $\vec{a}, \vec{b}$ in S and every $\lambda \in (0, 1)$,

$$f(\lambda\vec{a} + (1-\lambda)\vec{b}) \; < \; \lambda f(\vec{a}) + (1-\lambda) f(\vec{b}) \tag{1.19}$$

Then we say that f is *convex*. We describe f as *concave* iff $-f$ is convex. Equivalently, f is concave iff the condition (1.19) is always true with the inequality reversed. The following theorem is very helpful when we show that certain inference processes are well-defined.

Theorem 5 ([Egg]) *Let S be a convex subset of $\mathbb{D}^J$. Let $f : S \to \mathbb{R}$ be continuous and twice differentiable such that its second derivative is positive in all directions $\vec{u}$ parallel to S from all $\vec{x} \in S$. Then f is convex. If we also assume that V is a convex subset of S, then there exists a unique $\vec{x} \in V$ such that $f(\vec{v}) \geq f(\vec{x})$ for all $\vec{v} \in V$.*

1.2 Belief as probability

The philosophy of this work is that the degree of belief a rational agent has in a particular sentence θ being true should be viewed as the probability of θ in the view of that agent. In Subsections 1.2.1 and 1.2.2 we discuss justifications of this philosophy.

1.2.1 The urn model

The so-called *urn problem* is a probability problem in which balls are drawn at random from an urn, in which all of the balls are known to have one of a finite number of different colours. Given what we observe from a finite number of draws (where each ball is replaced after being drawn), we wish to infer the proportions of the different colours of balls in the urn. The idea dates back to the early 18th century work [Ber], and has also been studied by De Moivre and Bayes.

The following justification for belief as probability can be found in [[Par] pp. 17-18].

Suppose that a rational agent has formed beliefs regarding a natural phenomenon $\mathcal{P}$ which they have observed a large number of times, where each observation is denoted by X. This is analogous to the evidence gained when a large number of selections have been made from the urn. If we assume that the contents of the urn do not change over time and that each withdrawal is independent of all of the previous withdrawals, of which there are many, then the past frequency with which a particular colour of ball has been withdrawn is a statistically fair estimate of the (unknown) proportion of balls in the urn which have that colour.

If the natural phenomenon $\mathcal{P}$ is regarded in a similar way, i.e. we assume that there is an underlying natural frequency with which an occurrence of $\mathcal{P}$ satisfies a fixed sentence $\theta \in SL$, then the agent's belief in θ being true the next time they observe $\mathcal{P}$ is the proportion of times θ has been true in the past, i.e.

$$\frac{|\{X|X \models \theta\}|}{|\{X\}|} \tag{1.20}$$

Now we can prove (as in [[Par], p.19]) that this forces the agent's beliefs to be probabilistic. Indeed $Bel(\theta)$ is the probability that a previous occurrence of $\mathcal{P}$, chosen using the uniform distribution, satisfies θ.

Remark However, in practice, if we really have observed $\mathcal{P}$ a large number of times it is unreasonable that we remember every single one clearly. However, it can be argued that we may be subconsciously aware of all our previous memories when assessing, for example, based on the visible clouds, how likely it is to rain in the next hour. Also it is quite possible that the agent believes that ϕ is possible but has never observed it him/herself. Some past observations might be more indirect, such as those passed on from other observers and written in books or other media. The urn model, however, gives all observations equal weight. We often wish to infer beliefs about situations that have never happened before—something the urn model does not allow.

1.2.2 Belief as willingness to bet

The ideas of this subsection arise originally from work of [Ram], [deFin1], [deFin2], [Kem] and [Shi]. Our exposition of this argument here is inspired by [[Par], pp.19-22]. The extent to which the agent believes that θ is true can be regarded as the agent's willingness to bet that θ is true. We assume that the agent will place bets on whether sentences are true or false, before finding out for certain which atom α_{real} of the overlying language L is really true. We now specify the terms of the bets.

Suppose that for every $t \in [0, 1]$ and all $\theta \in SL$ the rational agent must choose between

(i) Gaining $(1 - t)$ if $\alpha_{real} \models \theta$ and losing t if $\alpha_{real} \models \neg\theta$

(ii) Losing $(1 - t)$ if $\alpha_{real} \models \theta$ and gaining t if $\alpha_{real} \models \neg\theta$

We now observe that, for all $\theta \in SL$, if $t = 0$ the agent must choose (i) as then he/she has nothing to lose. Similarly if $t = 1$, the agent must choose (ii).

Let $T', T \in [0, 1]$ such that $T' < T$.

Claim The agent is irrational if they choose (i) when $t = T$ and (ii) when $t = T'$, insofar as if, instead, they choose (i) when $t = T'$ and (ii) when $t = T$ they will be better off.

Proof of claim

Case 1 $\underline{\alpha_{real} \models \theta}$ Choosing (i) when $t = T$ and (ii) when $t = T'$ results in a gain of $1 - T$ and a loss of $1 - T'$ respectively. However choosing (i) when $t = T'$ and (ii) when $t = T$ results in a gain of $1 - T'$ and a loss of $1 - T$. The gain is now larger and the loss smaller.

Case 2 $\underline{\alpha_{real} \models \neg\theta}$ Choosing (i) when $t = T$ and (ii) when $t = T'$ results in a loss of T and a gain of T' respectively. However choosing (i) when $t = T'$ and (ii) when $t = T$ gives a loss of T' and a gain of T.

Hence whether θ turns out to be true or false in the real world, it is irrational to choose (i) for a larger value of t and (ii) for a smaller value. We have proved the claim.

$\blacksquare$

Thus for each $\theta \in SL$ there exists $\beta \in [0, 1]$ such that for all $t \in [0, 1]$ s.t. $t > \beta$ the agent chooses (ii) and for all $t < \beta$ their choice is (i). We now define this value to be the extent to which the agent is prepared to bet on θ being true since for $t > \beta$ it is more attractive to place the bet that pays out when $\alpha_{real} \models \neg\theta$ but if $t < \beta$ the agent prefers the bet that pays out when $\alpha_{real} \models \theta$.

Definition A *Dutch Book* against a belief function is a finite collection of bets on sentences $\theta \in SL$ such that, according to the rational agent's beliefs, they are compelled to accept the bets individually, but the agent loses overall no matter which $\alpha_{real} \in At^L$ represents the real world.

Definition If a rational agent forms their beliefs in terms of willingness to bet simultaneously for all $\theta \in SL$ so that they cannot change their mind, these beliefs are *fair* if and only if there does not exist a Dutch Book against these beliefs.

Theorem 6 *A belief function $Bel : SL \to [0, 1]$ is fair if and only if it is a probabilistic belief function.*

Proof This is similar to [[Par] p.21, p.23] apart from the fact that in [Par], the author defines the bets of type (i) and (ii) with variable stakes whereas in this work the stakes are fixed for each bet.

■

Remark Why should a rational agent be forced to bet? Indeed the author of [Chr] argues:

"*There is, after all, no Evil Super-bookie constantly monitoring everyone's credences, with an eye to making Dutch book against anyone who falls short of probabilistic perfection. Even if there were, many people would decline to bet at "fair odds", due to suspiciousness, or risk aversion, or religious scruples. In short, it is pretty clear that Dutch Book vulnerability is not, per se, a practical liability at all!*"

The author also outlines a method by which a Dutch book could exist against a rational agent, if we use a form of temporal logic and some of the sentences of the language refer to the beliefs the agent will hold tomorrow.

Whilst the "Evil Super-Bookie" does not exist, we do bet using our beliefs in everyday situations. For example, there is a certain "loss" incurred in carrying a cumbersome umbrella if it does not rain and a certain "loss" if we are rained upon because we do not carry the umbrella. Our belief in the statements "It will rain", "It will rain heavily" etc. determine which loss we risk incurring. If we view betting in this loose sense, the Dutch Book argument can become more credible.

One main philosophical point of view in choosing which inference processes to use is that we wish to maximise a measure of uncertainty to pick the point in $V^L(K)$ which is the least informative. We pursue this as a main theme in this thesis, although we also consider other approaches.

In the next section we introduce certain inference processes that have a great deal of importance.

1.3 The *Maximum Entropy* inference process

Definition For a language L and all $K \in CL$ the *Maximum Entropy* inference process, ME^L, is defined by

$$\begin{aligned} ME^L(K) = \text{ the unique } \vec{x} \in V^L(K) \text{ for which the Shannon} \\ \text{measure of uncertainty, } -\sum_{i=1}^{J} x_i \log(x_i) \text{ is maximal} \end{aligned} \tag{1.21}$$

where, by convention, $x \log(x) = 0$ when $x = 0$.

There are many justifications for using ME^L, including that of justifying the Shannon entropy as a measure of uncertainty. When we say that $-\sum_{i=1}^{J} x_i \log(x_i)$ is the measure of uncertainty of the p.b.f. $\vec{x}$, Shannon and Weaver provide the following justification in [Sha]:

We have a finite set of mutually exclusive events, one of which we know must occur, given by $\alpha_1, \alpha_2, \ldots \alpha_J$ where J might not equal 2^n with n the number of p.v.'s—we are just considering the outcomes so there may be any number of them. The belief values we have assigned to them are $Bel(\alpha_i) = x_i$ which are non-negative for each $i = 1, \ldots J$ and $\sum_{i=1}^{J} x_i = 1$. The amount of uncertainty is denoted by $H(\vec{x})$ and we think of that as the amount of information we expect to gain when we discover which event actually occurs.

The function

$$H : \cup_{J \geq 1} \mathbb{D}^J \to [0, \infty), \tag{1.22}$$

should satisfy the properties:

(a) For each $J \in \mathbb{N}$, $H \lceil \mathbb{D}^J$ (H restricted to $\mathbb{D}^J$) is continuous.

(b) For all integers J_1, J_2 s.t. $0 < J_1 < J_2$, $H\left(\frac{1}{J_1}, \ldots, \frac{1}{J_1}\right) < H\left(\frac{1}{J_2}, \ldots, \frac{1}{J_2}\right)$

(c) If $\sum_{j=1}^{M_i} y_{ij} = 1$ and $y_{i_j} \geq 0$ for $i = 1, 2, \ldots J$ and $\vec{x} \in \mathbb{D}^J$ then

$$\begin{aligned} &H(x_1 y_{1_1}, x_1 y_{1_2}, \ldots x_1 y_{1_{M_1}}, x_2 y_{2_1}, x_2 y_{2_2}, \ldots x_i y_{i_j} \ldots) \\ &\qquad = H(\vec{x}) + \sum_{i=1}^{J} x_i H(y_{i1}, y_{i2}, \ldots y_{iM_i}). \end{aligned} \tag{1.23}$$

We justify (a) because it can be expected that a small change in $\vec{x}$ should not significantly change $H(\vec{x})$. The level of uncertainty in J equally likely outcomes should increase as J increases, so we have (b). If you learn first which x_i event happens and then which y_{i_j} event happens the formula in (c) shows the expected amount of information gained, since it should make no difference whether the information comes all at once or in those two steps.

Theorem 7 ([Sha]) *Assume H is a function satisfying (a), (b) and (c) above. Then there exists a positive real constant c such that for all positive integers J and all $\vec{x} \in \mathbb{D}^J$,*

$$H(x_1, \ldots, x_J) = -\sum_{i=1}^{J} cx_i \log x_i \tag{1.24}$$

Proof The proof of the theorem as expressed above can be found in [[Par], pp. 77-78].

■

1.3.1 The "balls in boxes" justification of *Maximum Entropy*

The following derivation of *Maximum Entropy* is thought to have arisen from a suggestion by Graham Wallis to E.T. Jaynes in 1962—see [Jay]. Firstly, we write it in a non-rigorous form.

When trying to assign probabilities to a collection of mutually exclusive possible outcomes, the possible worlds $\alpha_1, \ldots \alpha_J$, imagine that the total probability of 1 comes as a large collection of N discrete amounts of $1/N$. The balls (quanta of $1/N$ probability) are now allocated at random to the boxes (the atoms) with total symmetry between the different balls and the different boxes. The balls land independently of each other. It is deemed that the fairest allocation of probability is that which is most likely to occur, out of those that satisfy our knowledge base.

Hand-waving "Proof" Consider the allocation of n_1 balls in box 1, n_2 in box 2, ... n_i in box i for each i s.t. $1 \leq i \leq J$ so that $Bel(\alpha_i) = x_i = n_i/N$. The probability that the balls land in that way is

$$\frac{N!}{n_1!n_2!\dots n_J!J^N} = P \tag{1.25}$$

When maximising P, we equivalently maximise $\frac{1}{N}\log_e(P.J^N)$ i.e.

$$\frac{1}{N}\log_e\left(\frac{N!}{n_1!n_2!\dots n_J!}\right) \tag{1.26}$$

$$= \frac{1}{N}\left(\log_e(N!) - \sum_{i=1}^{J}\log_e(Nx_i)!\right) \tag{1.27}$$

Now we use Stirling's approximation for the factorial function:

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}\ \left(\frac{n}{e}\right)^n} = 1 \tag{1.28}$$

We say $\approx$ means "approximately equals for large N" and assume that the large $N \to \infty$ so that (1.28) is used in the form

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{1.29}$$

so that

$$\log_e(n!) \approx c + (n + 1/2)\ \log_e\ n - n \tag{1.30}$$

where c is constant. Thus maximising P becomes equivalent, as $N \to \infty$, to maximising

$$\frac{1}{N}\left(\left(N+\frac{1}{2}\right)\log_e N - N - \sum_{i=1}^{J}\left(Nx_i+\frac{1}{2}\right)(\log_e N + \log_e x_i) + \sum_{i=1}^{J}Nx_i\right) \tag{1.31}$$

Expanding the above gives

$$\frac{1}{N}\left(\left(N+\frac{1}{2}\right)\log_e N - \frac{J\log_e N}{2} - N\log_e N - N\sum_{i=1}^{J}x_i\log_e x_i - \frac{1}{2}\sum_{i=1}^{J}\log_e x_i\right) \tag{1.32}$$

and the $N\log_e\ N$ terms cancel out, leaving

$$\log_e N\left(\frac{1-J}{2N}\right) - \sum_{i=1}^{J}\log_e x_i\left(x_i + \frac{1}{2N}\right) \tag{1.33}$$

Terms which are constant or small as $N \to \infty$ become immaterial in the limit, leaving us with the problem of maximising

$$-\sum_{i=1}^{J}x_i\log_e x_i \tag{1.34}$$

which is equivalent to us maximising the Shannon entropy, since $\log(x) = \log_e(x).\log(e)$ for all positive real values of x.

End of Hand-waving "Proof" Thus we have "proved" that we should maximise the Shannon entropy when N becomes very large.

The above argument is not a genuine mathematical derivation: however, in [ParVen3], the authors formulate and prove a rigorous theorem based on the same ideas, which we state overleaf.

When we assume that the rational agent's knowledge base arises from a large case history of examples, the convergence to ME^L can be seen to be more convincing than the above argument suggests. We reinterpret a constraint of the form

$$\sum_{j=1}^{r} a_{ji} Bel(\theta_j) = b_i \tag{1.35}$$

as

$$\sum_{j=1}^{r} a_{ji} |\{X \in M|\ X \text{ satisfies } \theta_j\}| = b_i |M| \tag{1.36}$$

where M is the large set of previous examples X from which the rational agent has gained the knowledge. Insisting on equality above puts severe conditions on the integer $|M| = N$. Hence the agent, slightly unsure of the exact numbers of previous experiments in which the θ_j are true, knows that

$$\sum_{j=1}^{r} a_{ji} |\{X \in M|\ X \text{ satisfies } \theta_j\}| \approx b_i N \tag{1.37}$$

which we make rigorous shortly. Now all that is known about M are the constraints of K, written in the form (1.37) and, to infer $Bel(\theta)$, we find an approximation for

$$\frac{|\{X \in M|X \text{ satisfies } \theta\}|}{N} \tag{1.38}$$

Considering all of the possible values of M that could have given rise to the knowledge base, it can be shown (see [ParVen1]) that ALMOST ALL of these, when M is large enough, infer $Bel(\theta) \approx ME^L(K)(\theta)$!

The notion of $\approx$ is now made more precise as a binary relation on $\{1, 2, \ldots N\}$. Where there are m constraints on sentences of n propositional variables with N fixed, for every $x, y \in \{1, \ldots N\}$,

$$\begin{aligned} |x-y| \leq 2^{n\sqrt{m}} \quad &\Rightarrow \quad x \approx y \\ x \approx y \quad &\Rightarrow \quad |x-y| \leq \epsilon N \end{aligned} \tag{1.39}$$

where ϵ is a fixed, positive real number.

Lemma 8 ([ParVen1]) *Fix a consistent knowledge base K. For every $\mu, \nu > 0$ there exist $\epsilon > 0$ and a positive integer N_0 such that for every $N \geq N_0$ and all relations $\approx$ satisfying the above conditions the proportion of possible M satisfying K for which*

$$\left| \frac{|\{X \in M | X \ satisfies \ \theta\}|}{N} - ME^L(K)(\theta) \right| \geq \nu \tag{1.40}$$

is less than or equal to μ.

In other words, the rational agent is compelled to take belief values closer and closer to those inferred by $Maximum\ Entropy$ when the size of previous evidence $N \to \infty$. In practice, even if the knowledge is obtained from historical frequencies we would not expect N to be large enough for the convergence to seem convincing. However, Lemma 8 is the first known rigorous version of the “Hand-waving Proof”, justifying using ME^L by the “balls in boxes” method.

1.4 The axiomatic approach—introducing the Par-Ven Properties

The inference processes we study in this thesis can mainly be justified by one or both of two approaches:

1 *Information theoretic* The inference process is defined as minimising a measure of information, so we justify choosing this particular measure.

2 *Axiomatic* We show that the inference process is the only one that satisfies all of a certain set of desirable properties, or desiderata.

Earlier in this introduction we have seen that the information theoretic approach can justify $Maximum\ Entropy$, using the Shannon entropy.

In [ParVen1], the authors use the axiomatic approach. They justify ME^L as the only inference process satisfying a list of properties or desiderata on a general inference process N^L. In Subsections 1.4.1-1.4.9 we define the nine Paris-Vencovska Properties, known as the Par-Ven Properties, and explain why we can think of them as "common-sense".

1.4.1 Equivalence Principle

For all knowledge bases $K_1, K_2 \in CL$ *such that* $V^L(K_1) = V^L(K_2)$,
$N^L(K_1) = N^L(K_2)$.

Definition If K_1, K_2 are consistent constraint sets on J co-ordinates and $V(K_1) = V(K_2)$, we say that K_1 and K_2 are *equivalent.* In this case we write $K_1 \equiv K_2$.

We now explain why Equivalence is a desirable property. If K_1 and K_2 are knowledge bases which impose the same conditions on p.b. functions then we should not care exactly how their constraints are expressed, so we should favour the same solution for both knowledge bases. The Watts Assumption helps us see that the knowledge contents of equivalent knowledge bases are essentially the same.

However, a possible criticism of this is that in practice it may sometimes be difficult for us to prove that K_1 and K_2 are equivalent. Perhaps we should merely insist that if we can prove in a straightforward way that $V^L(K_1) = V^L(K_2)$, then $N^L(K_1) = N^L(K_2)$.

This, though, is not a problem since the author in [[Par], pp.84–85] outlines a system of axioms and rules which can always show that $K_1 \equiv K_2$ when that is the case. In this thesis we only discuss inference processes that satisfy Equivalence.

1.4.2 Atomic Renaming Principle

Suppose that $K_1, K_2 \in CL$*,*

$$K_1 = \left\{ \sum_{j=1}^{J} a_{ji} Bel(\gamma_j) = b_i \,|\, i = 1, \ldots m \right\},$$

$$K_2 = \left\{ \sum_{j=1}^{J} a_{ji} Bel(\delta_j) = b_i \,|\, i = 1, \ldots m \right\},$$

where $\gamma_1, \ldots \gamma_J$*,* $\delta_1, \ldots \delta_J$ *are permutations of* $\alpha_1, \ldots \alpha_J$*. Then*

$$N^L(K_1)(\gamma_j) = N^L(K_2)(\delta_j)$$

for each $j = 1, \ldots J$*.*

This principle is saying that the atoms of the language have equal status as possible outcomes or worlds so if we permute them in the knowledge base then the beliefs we infer should be permuted in the corresponding way. This can be challenged on the grounds that if it is accepted that permuting the propositional variables or reversing true and false for p.v.'s should produce symmetrical permutation of the beliefs, that of itself does *not* imply Atomic Renaming and so the latter principle may be unnecessarily strong.

Atomic Renaming means that it does not matter how we have labelled the atoms $\alpha_1, \ldots \alpha_J$, as long as we do this in a consistent manner. Apart from in Subsection 1.4.9, in this work we only consider inference processes that satisfy Atomic Renaming.

1.4.3 Obstinacy Principle

Suppose $K_1, K_2 \in CL$ and $N^L(K_1)$ satisfies K_2. Then $N^L(K_1 + K_2) = N^L(K_1)$.

This is justified by observing that if the $N^L(K_1)$ satisfies K_2 then on the basis of K_1 the inference process is telling you to infer K_2. Now if we believe K_2 as well as K_1 before applying the inference process the same beliefs should be chosen. Notice that Equivalence is a consequence of Obstinacy, if we are not already regarding Equivalence as a given:

Assume that K_1, K_2 are equivalent and N^L satisfies Obstinacy. Then $N^L(K_1)$ satisfies K_1 so it satisfies K_2. Hence $N^L(K_1 + K_2) = N^L(K_1)$. Similarly $N^L(K_2 + K_1) = N^L(K_2)$. Now $K_1 + K_2$ is the *same* knowledge base as $K_2 + K_1$, so $N^L(K_1) = N^L(K_2)$ and Equivalence is satisfied.

Theorem 9 *If an inference process N^L is given by:*

$$N^L(K) = \text{ the unique } \vec{x} \in V^L(K) \text{ which is minimal w.r.t. } \hat{\leq} \tag{1.41}$$

where $\hat{\leq}$ is a partial ordering on $\mathbb{D}^J$, N^L satisfies Obstinacy.

Proof Suppose that N^L takes the above form and that $\vec{X} = N^L(K_1) \in V^L(K_2)$. Then $\vec{X} \in V^L(K_1 + K_2)$. If there exists $\vec{Y} \in V^L(K_1 + K_2)$ such that $\vec{Y} \hat{\leq} \vec{X}$, then $\vec{Y} \in V^L(K_1)$, which is a contradiction. Hence $\vec{X} = N^L(K_1 + K_2)$, so N^L satisfies Obstinacy and we have proved the theorem.

■

The following remark, due to [[Par], page 99], is similar to Theorem 9.

Remark An inference process N^L is Obstinate if it can be defined as:

$$\text{For all } K \in CL, (N^L)(K) = \text{ the unique } \vec{x} \in V^L(K) \text{ such that } F(\vec{x}) \text{ is minimal} \tag{1.42}$$

where $F : \mathbb{D}^J \to Q$, for a totally ordered set Q.

1.4.4 Language Invariance

Suppose we have a family of inference processes N^L, one for each finite language L. Then this family is said to be Language Invariant *if whenever $L_1 \subseteq L_2$ (so that $SL_1 \subseteq SL_2$ and $CL_1 \subseteq CL_2$) and $K \in CL_1$, then $N^{L_2}(K)$ agrees with $N^{L_1}(K)$ on SL_1.*

This is a unifying property of the family of inference processes. We justify this by supposing that if K only refers to $p_1, p_2, \ldots p_n$, then, given that we are inferring beliefs for sentences only mentioning $p_1, \ldots p_n$, it should make no difference whether p_{n+1} is in the overlying language or not.

1.4.5 Principle of Continuity

Intuitively it is desirable that an inference process be continuous, i.e. that a microscopic change in information should not produce a macroscopic change in inferences. It can be argued that the knowledge held by the rational agent is possibly fluctuating a little and it would be unreasonable if these variations produced significant changes in the beliefs inferred. The difficulty in stating this property is to find a suitable topology on knowledge bases, to answer the question "When are knowledge bases close to each other?"

An obvious first attempt is to say that two knowledge bases, K_1 and K_2, are close iff their coefficients in the constraints are close. Let them have respective matrices A and B of coefficients, given by $A = (a_{ji})$, $B = (b_{ji})$ and let

$$||A - B|| = max|a_{ji} - b_{ji}| \tag{1.43}$$

However, this does not do the job as there are always knowledge bases equivalent to K_1 with a different matrix of coefficients, say $C = c_{ji}$, and if a knowledge base had a matrix D close to C it wouldn't be close to A. Thus our definition of closeness must take knowledge content into account.

Definition The *Blaschke metric* measures distance between convex subsets C

and D of $\mathbb{D}^J$ by

$$\Delta(C,D) = inf\{\delta \text{ s.t. for all } \vec{x} \in C \text{ there exists } \vec{y} \in D, \text{ s.t. } |\vec{x}-\vec{y}| \leq \delta$$
$$\text{and for all } \vec{y} \in D \text{ there exists } \vec{x} \in C, \text{ s.t. } |\vec{x}-\vec{y}| \leq \delta\} \quad (1.44)$$

where $|\vec{x}-\vec{y}|$ is the usual Euclidean distance between vectors.

Remark Why should we use *Euclidean* distance between the vectors representing p.b.f.'s, instead of any other measure of distance between them?

One commonly used measure is the *Kullback-Liebler divergence*, which is also known as the cross-entropy distance between two p.b.f.'s. We discuss this measure in Subsection 1.5.2. and it is given by (1.70) in the case that $\vec{x}$ and $\vec{y}$ are p.b.f.'s (i.e. $W(\vec{x}) = W(\vec{y}) = 1$, where $\vec{x}$ and $\vec{y}$ are finite discrete generalised probability distributions the sense of (1.58)) and for each $i = 1, \ldots J$, $x_i = 0$ implies that $y_i = 0$:

$$D_{K,L}(\vec{y},\vec{x}) = \mathcal{I}_1(\vec{y}|\vec{x}) = \sum_{i=1}^{J} y_i \log\left(\frac{y_i}{x_i}\right) \quad (1.45)$$

where we let $x_i/y_i = 1$ if $x_i = y_i = 0$. It represents the amount of information we gain when our observance of a phenomenon causes us to change our probabilistic beliefs from $\vec{x}$ to $\vec{y}$. For a full justification of this (derived from [Ren1]), see Subsection 1.5.2.. However, Kullback-Liebler divergence is unsuitable for constructing definitions of continuity of inference processes etc.. because it is not a metric on p.b.f.'s as it is not symmetric and does not satisfy the triangle inequality.

Definition The *Blaschke topology* on the set of equivalence classes of CL is the topology induced by the Blaschke metric on the set of solution sets $V^L(K)$. In general, we shall write $\Delta(K_1, K_2)$ in place of $\Delta(V^L(K_1), V^L(K_2))$. Thus N^L satisfies *Continuity* iff $N^L : CL \to \mathbb{D}^J$ is a continuous function, using the Blaschke topology on the equivalence classes of CL and the Euclidean topology on $\mathbb{D}^J$.

We can use this measure of distance to take limits of some sequences of knowledge bases. For example, if $K_\epsilon \in CL$ for all $\epsilon \in [0,\delta]$ and some $\delta > 0$, we may write

$$K_\epsilon \to K_0 \text{ as } \epsilon \searrow 0 \quad (1.46)$$

to mean that $\Delta(K_0, K_\epsilon) \to 0$ as $\epsilon \searrow 0$.

Definition For bounded non-empty convex sets A, $B \in \mathbb{R}^J$ define $\Delta_{A\to B} = inf\{\epsilon$ s.t. for all $\vec{x} \in A$ there exists $\vec{y} \in B$ s.t. $|\vec{y} - \vec{x}| \le \epsilon\}$.

Lemma 10 *For all such A, B, $\Delta(A, B) = max(\Delta_{A\to B}, \Delta_{B\to A})$.*

Proof The set $C = \{\epsilon$ s.t. for all $\vec{x} \in A$ there exists $\vec{y} \in B$ s.t. $|\vec{y} - \vec{x}| \le \epsilon\}$ is such that, for each $z \in \mathbb{R}$, $z \in C \Rightarrow [z, \infty) \subseteq C$. Hence if $\Delta_{A\to B} \in C$, $C = [\Delta_{A\to B}, \infty)$. Otherwise for all $n \in \mathbb{N}$ there exist members of C, c_n, such that

$$\Delta_{A\to B} < c_n < \Delta_{A\to B} + 1/n \tag{1.47}$$

so, since

$$\bigcup_{n=1}^{\infty} (\Delta_{A\to B} + 1/n, \infty) = (\Delta_{A\to B}, \infty) \subseteq C \tag{1.48}$$

$C = (\Delta_{A\to B}, \infty)$ in this case.

Now we consider

$$\begin{aligned} E = \{\epsilon \text{ s.t. for all } \vec{x} \in A \text{ there exists } \vec{y} \in B \text{ s.t. } |\vec{y} - \vec{x}| \le \epsilon \\ \text{and for all } \vec{x} \in B \text{ there exists } \vec{y} \in A \text{ s.t. } |\vec{y} - \vec{x}| \le \epsilon\} \end{aligned} \tag{1.49}$$

Then E is the intersection of a set with infimum $\Delta_{A\to B}$ and one with infimum $\Delta_{B\to A}$ so its infimum is $\Delta(A, B) = max(\Delta_{A\to B}, \Delta_{B\to A})$ and we have proved the lemma. ■

Notation For all $\vec{a}, \vec{b} \in \mathbb{R}^J$, let $[\vec{a}, \vec{b}]$ denote the line segment which has end points $\vec{a}$, $\vec{b}$.

Lemma 11 *Let $\vec{a}_i$, $\vec{b}_i$ be vectors in $\mathbb{R}^J$, for $i = 1, 2$. Where ϵ is a positive real number, if $|\vec{a}_1 - \vec{a}_2| < \epsilon$ and $|\vec{b}_1 - \vec{b}_2| < \epsilon$ then*

$$\Delta([\vec{a}_1, \vec{b}_1], [\vec{a}_2, \vec{b}_2]) < \epsilon \tag{1.50}$$

Proof For a general point of $[\vec{a}_1, \vec{b}_1]$, say $\mu\vec{a}_1 + (1 - \mu)\vec{b}_1$, its distance from the point $\mu\vec{a}_2 + (1 - \mu)\vec{b}_2$ is less than $\mu\epsilon + (1 - \mu)\epsilon$. Similarly, for every point in $[\vec{a}_2, \vec{b}_2]$ there exists a point in $[\vec{a}_1, \vec{b}_1]$ such that the distance between them is less than ϵ. Hence we have proved the lemma.

■

Theorem 12 ([Mau]) *If f is a convex continuous function and $K \in CL$, then there exists a unique $\vec{x} \in V^L(K)$ such that $f(\vec{x})$ is minimal.*

Theorem 13 ([Mau]) *If we define an inference process N^L by:*

For all $K \in CL$, $N^L(K) =$ the unique $\vec{x} \in V^L(K)$ such that $f(\vec{x})$ is minimal

where f is a fixed continuous convex function, then N^L is well-defined and continuous.

1.4.6 Open-mindedness Principle

If $K \in CL$, $\theta \in SL$ and $K + Bel(\theta) = c$ is consistent for some $c > 0$
then $N^L(K)(\theta) > 0$

The justification for this is that if our knowledge does not force us into $Bel(\theta) = 0$ then we should not infer that belief as that would be to unnecessarily condemn θ as definitely not true. There is an argument against this, namely that there is a difference between $\models \neg\theta$ and $Bel(\theta) = 0$, particularly in a predicate context—if every rational number between 1 and 2 is equally likely (a priori) to be chosen by a procedure we should give every such possible outcome belief 0 but they are all logically possible.

However, in real life we may often, in practice, ignore small probabilities, such as that of being struck by lightning. If we accept that $Bel(\theta) = 0$ is an extreme view then Open-mindedness is a desirable property of an inference process.

1.4.7 Principle of Independence

In the special case of $L = \{p_1, p_2, p_3\}$ and

$$K = \{Bel(p_1) = a, Bel(p_2|p_1) = b, Bel(p_3|p_1) = c\} \textit{ where } a > 0,$$

$$N^L(K)(p_2 \wedge p_3|p_1) = bc.$$

The justification for this principle is that since K provides no connection between the propositional variables p_2 and p_3 on the basis of p_1, they should be treated as being statistically independent. It is possible, though, to have slightly different ideas of independence and we can certainly question how high the status of this notion of independence should be.

1.4.8 Relativisation Principle

Suppose that $K_1, K_2 \in CL, 0 < c < 1$ *and*

$$K_1 = \{Bel(\phi) = c\} + \left\{ \sum_{j=1}^{r} a_{ji} Bel(\theta_j|\phi) = b_i \,|\, i = 1, \dots m \right\},$$

$$K_2 = K_1 + \left\{ \sum_{j=1}^{q} e_{ji} Bel(\psi_j|\neg\phi) = f_i \,|\, i = 1, \dots s \right\}.$$

Then for $\theta \in SL$, $N^L(K_1)(\theta|\phi) = N^L(K_2)(\theta|\phi)$.

This principle can seem sensible because, given that ϕ is true, the constraints of K_2 are no more than those of K_1 for $Bel(\theta|\phi)$; adding conditions on the values of $Bel(\theta|\neg\phi)$ should have no effect.

1.4.9 Principle of Irrelevant Information

Suppose that $K_1, K_2 \in CL$, $\theta \in SL$ *but that no propositional variable appearing in* θ *or in some sentence of* K_1 *also appears in a sentence in* K_2. *Then*

$$N^L(K_1 + K_2)(\theta) = N^L(K_1)(\theta) \tag{1.51}$$

This property can be attractive because the knowledge of K_2 seems to be irrelevant to K_1 and θ, since it involves completely different propositional variables, so once we know K_1 our belief in θ should be the same as if we also know K_2. It is proved on page 88 of [Par] that, whenever K_1 and K_2 are consistent and mention totally different propositional variables, $K_1 + K_2$ is consistent.

Remark Prior to the present work, ME^L was the only inference process known to satisfy Irrelevant Information and Atomic Renaming. The so-called *Observant*

Process (see [Court]), denoted by OP, is also known to satisfy Irrelevant Information, though it is not a single-valued function and so not an inference process. To be precise, when K_1, K_2, θ, L are as in the definition of Irrelevant Information,

$$OP(K_1 + K_2)(\theta) = OP(K_1)(\theta) \tag{1.52}$$

where equality means that the *sets* of possible values of $Bel(\theta)$, where Bel is in the set $OP(K..)$ are equal. We can use *Maximum Entropy* on the set of p.b.f.'s given by OP to reduce our inference to a single p.b.f.—this gives a true inference process which satisfies Irrelevant Information but not Atomic Renaming.

The following characterisations are due to the authors in [ParVen1]:

Theorem 14 ([ParVen1]) *If the only knowledge bases permitted were those using just rational number coefficients, ME^L is the only Language Invariant inference process satisfying the principles of Equivalence, Open-mindedness, Atomic Renaming, Obstinacy, Relativisation, Independence and Irrelevant Information.*

Remark Since ME^L is Language Invariant, we shall usually refer to it as ME from now on, not mentioning the overlying language. Furthermore, using the density of $\mathbb{Q}$ in $\mathbb{R}$, we can remove the assumption of Open-mindedness:

Theorem 15 ([ParVen1]) *ME is the only Language Invariant inference process which is continuous and satisfies Equivalence, Atomic Renaming, Obstinacy, Relativisation, Independence and Irrelevant Information.*

1.4.10 Representation dependence of inference processes

It is possible to criticise *Maximum Entropy* as "representation dependent" in the sense that it fails the following desideratum. However, this view does not appear to be coherent.

Atomicity Principle: *Let $\theta \in SL_2$ be neither a contradiction nor a tautology, $K \in CL_1$, $\phi \in SL_1$, $L_1 \cap L_2 = \emptyset$, $L_1, L_2 \subseteq L$. Let ϕ^θ etc. be the result of replacing a particular propositional variable $p \in L_1$ everywhere by θ. Then*

$N^L(K)(\phi) = N^L(K^\theta)(\phi^\theta)$.

This principle would be justified by supposing that propositional variables are merely where the detail of the statement of the inference problem stops and that nothing should change if, when we inspect the world more closely, we turn a propositional variable into a sentence composed of new p.v.'s.

Theorem 16 ([Par], p.102) *No inference process satisfies the Atomicity Principle.*

Although Atomicity can seem very reasonable, it is in fact contradictory! Furthermore, in [HalKol] the authors show that, even when inference processes are permitted to select a set of solutions of the knowledge base, rather than just one solution, no such process can be representation independent. Arguments refuting the criticism that *Maximum Entropy* is representation dependent, and more, can be found in [ParVen4].

1.5 Introducing the *Renyi Processes*

A major motivation for this thesis is the comparison of other inferences processes with *Maximum Entropy* in terms of the Par-Ven Properties.

The *Renyi Processes* are named after the mathematician Alfred Renyi, who worked on them in the mid-20th century—see [Ren1] and [Ren2]. The author of [Uff] shows that the axioms used by the authors in [ShoJoh] can be restated mathematically to justify using the *Renyi Processes*.

In 1998 the author of [Moh] considered these inference processes, which he called the "Renyi generalized entropies in the discrete case".

Definition For each positive real number r such that $r \neq 1$ define the *Renyi Entropy* with parameter r to be H_r, given by

$$H_r(\vec{x}) = -\frac{\log(\sum_{i=1}^{J} x_i^r)}{r-1} \tag{1.53}$$

for all p.b.f.'s $\vec{x}$ over a language L s.t. $|At^L| = J$. The *Renyi Process* with parameter r is Ren_r^L, given by

$$Ren_r^L(K) = \text{ the unique } \vec{x} \in V^L(K) \text{ for which } H_r(\vec{x}) \text{ is maximal} \tag{1.54}$$

for all $K \in CL$. Note that in [Moh], the author parameterises each *Renyi Entropy* and *Renyi Process* using values of r that are 1 less than in the notation we will use in this thesis, as defined above. We now reexpress the *Renyi Processes* in a more concise form. If $r > 1$,

$$Ren_r^L(K) = \text{ the unique } \vec{x} \in V^L(K) \text{ for which } \sum_{i=1}^{J} x_i^r \text{ is minimal} \tag{1.55}$$

and, if $0 < r < 1$,

$$Ren_r^L(K) = \text{ the unique } \vec{x} \in V^L(K) \text{ for which } \sum_{i=1}^{J} x_i^r \text{ is maximal} \tag{1.56}$$

As yet Ren_1^L is undefined. However, the following definition implies that the Ren_r^L form a continuum as r varies through the positive real numbers.

Definition For any $\vec{x} \in \mathbb{D}^J$, $H_1(\vec{x}) = -\sum_{i=1}^J x_i \log(x_i)$, i.e. the Shannon entropy. The inference process Ren_1^L is defined to be *Maximum Entropy*.

In [Moh], the author shows that for every p.b.f. $\vec{x}$, $H_r(\vec{x})$ is a continuous function of r. Also, for any $K \in CL$ and $r_0 \in \mathbb{R}$ such that $r_0 > 0$,

$$Ren_r^L(K) \to Ren_{r_0}^L(K) \text{ as } r \to r_0 \tag{1.57}$$

It is shown in [Moh] that the *Renyi Processes* all satisfy Equivalence, Atomic Renaming, Obstinacy, Language Invariance, Continuity and Relativisation. From now on we shall usually refer to Ren_r without mentioning the overlying language. Ren_r satisfies Open-mindedness iff $r \leq 1$ and none of the *Renyi Processes* Ren_r for which $r \neq 1$ satisfy Independence. It was also shown in [Moh] that none of the Ren_r for which $r > 1$ satisfy Irrelevant Information and this property was conjectured to fail for Ren_r when $r < 1$.

In the following two subsections we detail two arguments, by Renyi himself, that provide justification for using the *Renyi Processes*. In [Ren1], the author first justifies the Shannon entropy as the only uncertainty measure that satisfies a list of five postulates that are attributed to Fadeev, see [Fad].

Notation We use different notation to Renyi, for the consistency of this thesis. However we stick closely to the content of [Ren1].

For any $J \in \mathbb{N}$, define

$$\mathbb{D}*^J = \left\{ \vec{x} \in \mathbb{R}^J \text{ s.t. } x_i \geq 0 \text{ for each } i = 1, \dots J \text{ and } 0 < \sum_{i=1}^J x_i \leq 1 \right\} \tag{1.58}$$

The members of $\mathbb{D}*^J$ are known as *finite discrete generalised probability distributions* and the *weight* of each $\vec{x} \in \mathbb{D}*^J$ is denoted by $W(\vec{x})$, and is given by $W(\vec{x}) = \sum_{i=1}^J x_i$.

1.5.1 Renyi's five postulates and their consequences

We consider an entropy function

$$H : \bigcup_{J \geq 1} \mathbb{D}*^J \to \mathbb{R} \tag{1.59}$$

and Renyi's ([Ren1]) list of postulates is:

Postulate 1: *$H(\vec{x})$ is a symmetric function of the co-ordinates.*

We can justify this on the same basis as Atomic Renaming.

Postulate 2: *$H((p))$ is a continuous function of p as p varies in $(0,1]$.*

It can seem unreasonable if a microscopic change in p causes a macroscopic change in $H((p))$.

Postulate 3: *$H((1/2)) = 1$.*

This postulate merely normalises the function H.

Definition If $\vec{x} \in \mathbb{D}*^J$ and $\vec{y} \in \mathbb{D}*^Q$ then we define

$$(\vec{x} \otimes \vec{y}) = (x_1y_1, x_1y_2, \ldots x_1y_Q, x_2y_1, \ldots x_2y_Q \ldots\ldots x_Jy_1, x_Jy_2, \ldots x_Jy_Q) \in \mathbb{D}*^{JQ} \tag{1.60}$$

and note that $W(\vec{x} \otimes \vec{y}) = W(\vec{x})W(\vec{y})$.

Also if $\vec{x} \in \mathbb{D}*^J$ and $\vec{y} \in \mathbb{D}*^Q$ such that $W(\vec{x}) + W(\vec{y}) \leq 1$ we define $\vec{x} \cup \vec{y} = (x_1, x_2, \ldots x_J, y_1, y_2, \ldots y_Q)$, which is known as the *union* of $\vec{x}$ and $\vec{y}$.

Postulate 4: *If $\vec{x} \in \mathbb{D}*^J$ and $\vec{y} \in \mathbb{D}*^Q$ then*

$$H(\vec{x} \otimes \vec{y}) = H(\vec{x}) + H(\vec{y}) \tag{1.61}$$

We can justify this postulate similarly to condition (c) of the Shannon/Weaver justification for *Maximum Entropy*—see (1.23).

Postulate 5: *If $\vec{x} \in \mathbb{D}*^J$ and $\vec{y} \in \mathbb{D}*^Q$ are such that $W(\vec{x}) + W(\vec{y}) \leq 1$ then*

$$H(\vec{x} \cup \vec{y}) = \frac{W(\vec{x})H(\vec{x}) + W(\vec{y})H(\vec{y})}{W(\vec{x}) + W(\vec{y})} \tag{1.62}$$

This postulate tells us that the entropy of the union of two distributions is the weighted average of their entropies, with each entropy value weighted by the weight of the original distributions.

Theorem 17 ([Ren1]) *If H is a function satisfying Postulates 1,2,3,4,5 as above, H must be the function H_1 given by*

$$H_1(\vec{x}) = -\frac{\sum_{i=1}^J x_i \log x_i}{\sum_{i=1}^J x_i} \tag{1.63}$$

for all $\vec{x} \in \mathbb{D}{*}^J$.

Now, following [Ren1], we weaken the list of five postulates to characterise the measures of uncertainty H_r. This is done by replacing Postulate 5 with:

Postulate 5: There exists a strictly monotonic and continuous function* $g : \mathbb{R}^+ \to \mathbb{R}^+$ *such that if* $\vec{x} \in \mathbb{D}{*}^J$ *and* $\vec{y} \in \mathbb{D}{*}^Q$ *are such that* $W(\vec{x}) + W(\vec{y}) \leq 1$ *then*

$$H(\vec{x} \cup \vec{y}) = g^{-1}\left[\frac{W(\vec{x})g(H(\vec{x})) + W(\vec{y})g(H(\vec{y}))}{W(\vec{x}) + W(\vec{y})}\right] \tag{1.64}$$

Note that the conditions on g imply that the right hand side of (1.64) is well-defined. One possibility for the weighting function g which is compatible with Postulate 4 is that it be an exponential function. We can let the function g be given by

$$g_r(x) = 2^{(1-r)x} \tag{1.65}$$

for all x, where r is a positive real number not equal to 1.

Theorem 18 ([Ren1]) *If* H *is a function satisfying Postulates 1,2,3,4, and 5* and* r *is a fixed positive real number not equal to 1 such that for Postulate 5** $g = g_r$, *(as defined by* (1.65)*) then for all* $J \geq 1$ *and* $\vec{x} \in \mathbb{D}{*}^J$

$$H(\vec{x}) = H_r(\vec{x}) = -\frac{1}{r-1}\log\frac{\sum_{i=1}^{J} x_i^r}{\sum_{i=1}^{J} x_i} \tag{1.66}$$

Remark We remark, however, that it is far from clear why we should choose g to be an exponential function, as above, when the identity function, amongst others, is also consistent with Postulate 4 and allows us to uniquely characterise $Maximum$ $Entropy$.

1.5.2 An alternative approach by Renyi

Renyi further improves ([Ren1]) on the justification above for the *Renyi Processes* as follows:

Suppose that $\vec{x}$ represents the probability distribution across a selection of mutually exclusive outcomes, numbered $1, \ldots J$, which may occur as a result of some experiment we carry out on a natural phenomenon. Not all of the outcomes are necessarily referred to by $\vec{x}$—we use *generalised* probability distributions, as in Subsection 1.5.1.

If we now observe the event E, which might be a relevant property of the phenomenon we are investigating, the probabilities of the corresponding outcomes change from x_i to y_i for each $i = 1, \ldots J$. We assume that the information gained from the outcome of the experiment depends only on $\vec{x}$ and $\vec{y}$. Note that if $x_i = 0$, then $y_i = 0$ as long as we assume that if our belief in an outcome is zero, then that outcome is impossible. The information we gain about the experiment's outcome, as a result of us observing E, is denoted by

$$\mathcal{I}(\vec{y}|\vec{x}) \tag{1.67}$$

As before we will let $\vec{x}, \vec{y}$ be *generalised* probability distributions. We suppose that for all $J \in \mathbb{N}$ and $\vec{x}, \vec{y} \in \mathbb{D}*^J$ such that if $x_i = 0$ for some $i = 1, \ldots J$, then $y_i = 0$, the value $\mathcal{I}(\vec{y}|\vec{x})$ is defined such that the following Postulates are satisfied:

Postulate 6: *$\mathcal{I}(\vec{y}|\vec{x})$ is unchanged if the same permutation is used to rearrange the values of both $\vec{x}$ and $\vec{y}$.*

We justify this postulate by seeing that a relabelling of the outcomes does not essentially change our knowledge.

Postulate 7: *If $x_i \leq y_i$ for each $i = 1, \ldots J$ then $\mathcal{I}(\vec{y}|\vec{x}) \geq 0$. If $y_i \leq x_i$ for each $i = 1, \ldots J$ then $\mathcal{I}(\vec{y}|\vec{x}) \leq 0$.*

It can seem reasonable that if observing the event E does not reduce any of the probabilities of the outcomes listed by $\vec{x}$, then we do not have less information than before. Similarly, if none of the probabilities have increased, we have not gained any information.

Postulate 8: *$\mathcal{I}((1)|(\frac{1}{2})) = 1$.*

This postulate merely normalises the information function.

Postulate 9: *Suppose that $\mathcal{I}(\vec{y_1}|\vec{x_1})$ and $\mathcal{I}(\vec{y_2}|\vec{x_2})$ are both defined and that $\vec{X} = \vec{x}_1 \otimes \vec{x}_2$ and $\vec{Y} = \vec{y}_1 \otimes \vec{y}_2$ such that the enumeration of the outcomes for $\vec{X}$ and*

$\vec{Y}$ are as in the definition of $\otimes$. Then

$$\mathcal{I}(\vec{Y}|\vec{X}) = \mathcal{I}(\vec{y}_1|\vec{x}_1) + \mathcal{I}(\vec{y}_2|\vec{x}_2) \tag{1.68}$$

We can justify this postulate by considering that $\vec{x}_1$ and $\vec{y}_1$ refer to a totally unrelated experiment to $\vec{x}_2$ and $\vec{y}_2$ so the quantities of information gained can either be calculated separately and then added together, or we can think of the two experiments together as one, giving $\mathcal{I}(\vec{Y}|\vec{X})$.

Postulate 10: *There exists a continuous and strictly increasing function g defined on all real numbers whose inverse we denote by g^{-1} (where it is defined) such that if $\mathcal{I}(\vec{y}_1|\vec{x}_1)$ and $\mathcal{I}(\vec{y}_2|\vec{x}_2)$ are defined and $0 < W(\vec{x}_1) + W(\vec{x}_2) \leq 1$ and $0 < W(\vec{y}_1) + W(\vec{y}_2) \leq 1$ and the enumerations of the outcomes of $\vec{x}_1 \cup \vec{x}_2$ and of $\vec{y}_1 \cup \vec{y}_2$ are as in the definition of union then*

$$\mathcal{I}(\vec{y}_1 \cup \vec{y}_2|\vec{x}_1 \cup \vec{x}_2) = g^{-1}\left\{\frac{W(\vec{y}_1)g(\mathcal{I}(\vec{y}_1|\vec{x}_1)) + W(\vec{y}_2)g(\mathcal{I}(\vec{y}_2|\vec{x}_2))}{W(\vec{y}_1) + W(\vec{y}_2)}\right\} \tag{1.69}$$

Note that the above expression is a weighted average of two values taken by the function g, so the right hand side is well-defined because g is strictly increasing.

The following theorem shows that we do not need to make an arbitrary choice for the function g:

Theorem 19 ([Ren1]) *Suppose that the quantity $\mathcal{I}(\vec{y}|\vec{x})$ is defined for all $J \in \mathbb{N}$ and all $\vec{x}$, $\vec{y} \in \mathbb{D}*^J$ such that if $x_i = 0$ for some $i = 1, \dots J$, then $y_i = 0$. Suppose also that $\mathcal{I}$ satisfies Postulates 6,7,8,9 and 10. Then the function g in Postulate 10 must be either a linear or exponential function. In the first case $\mathcal{I}(\vec{y}|\vec{x}) = \mathcal{I}_1(\vec{y}|\vec{x})$ for all $\vec{x}$, $\vec{y} \in \mathbb{D}*^J$, where*

$$\mathcal{I}_1(\vec{y}|\vec{x}) = \frac{\sum_{i=1}^J y_i \log \frac{y_i}{x_i}}{\sum_{i=1}^J y_i} \tag{1.70}$$

and we define $y_i/x_i = 1$ if $x_i = y_i = 0$. In the second case $\mathcal{I}(\vec{y}|\vec{x}) = \mathcal{I}_r(\vec{y}|\vec{x})$ for all $\vec{x}$, $\vec{y} \in \mathbb{D}*^J$ and some positive real $r \neq 1$, where

$$\mathcal{I}_r(\vec{y}|\vec{x}) = \frac{1}{r-1} \log \frac{\sum_{i=1}^J \frac{y_i^r}{x_i^{r-1}}}{\sum_{i=1}^J y_i} \tag{1.71}$$

Now we can make inferences in the manner of an inference process and we can assume that we have no initial information about the likelihood of the outcomes $1, \ldots J$ of the experiment, which are exhaustive. Our initial probability distribution, denoted by $\vec{x}$ in the above notation, is $(\frac{1}{J}, \frac{1}{J}, \ldots \frac{1}{J})$, by symmetry. Then we wish to minimise the amount of information we have gained by changing to the distribution $\vec{y}$ after observing event E. Substituting $\frac{1}{J}$ for x_i for each $i = 1, \ldots J$ in the $\mathcal{I}_r$ formulae of the statement of Theorem 19 gives:

$$\mathcal{I}_1(\vec{y}|\vec{x}) = \frac{\sum_{i=1}^{J} y_i \log(Jy_i)}{\sum_{i=1}^{J} y_i} = \sum_{i=1}^{J} y_i \log y_i + \sum_{i=1}^{J} y_i \log J = -H_1(\vec{y}) + \log J \quad (1.72)$$

so if we choose the $\vec{y} \in V^L(K)$ minimising this quantity we are using *Maximum Entropy*.

Similarly, if r is positive, real and not equal to 1, the substitution of $\frac{1}{J}$ for each x_i gives

$$\mathcal{I}_r(\vec{y}|\vec{x}) = \frac{1}{r-1} \log \left(\sum_{i=1}^{J} (J^{r-1}) y_i^r \right) = \log J + \frac{1}{r-1} \log \sum_{i=1}^{J} y_i^r = -H_r(\vec{y}) + \log J \quad (1.73)$$

so if we choose the $\vec{y} \in V^L(K)$ minimising this quantity we are using Ren_r.

Remark The approach of this subsection can be seen to be a more convincing justification of the *Renyi Processes*, compared to the previous argument. Indeed, our choice of weighting function g for Postulate 10 turns out to be forced (by Theorem 19) whereas, in Postulate 5, an arbitrary choice of g was made.

1.5.3 To justify *Renyi Processes* with integer parameter

Here is another justification for using Ren_r when r is an integer greater than 1.

Suppose that in a laboratory a scientist is conducting an experiment in which the result, one of J mutually exclusive outcomes, is known to be random given fixed initial conditions. He/she is trying to determine the natural probabilities of the outcomes by repeating the experiment. It is reasonable to assume that the repetitions are independent and that the scientist would find it surprising if the same outcome happened every time. For r trials, where $r > 1$, we can minimise our belief in that occurrence, i.e. minimise $\sum_{i=1}^{J} x_i^r$. This means that we use Ren_r.

1.5.4 The *Minimum Distance* inference process

Definition ([ParVen2]) The *Minimum Distance* inference process is defined by:

$$\begin{aligned} MD^L(K) &= \text{the nearest point in } V^L(K) \text{ to } \left(\frac{1}{J}, \frac{1}{J} \dots \frac{1}{J}\right) \in \mathbb{D}^J \\ &\quad\ \text{using the Euclidean metric} \\ &= \text{the unique } \vec{x} \in V^L(K) \text{ for which } \sum_{i=1}^{J} (x_i - 1/J)^2 \text{ is minimal.} \end{aligned} \tag{1.74}$$

We can justify using this inference process by the argument that $\vec{X} = (\frac{1}{J}, \dots \frac{1}{J})$ is the fairest p.b. function to infer when we have no knowledge at all, by symmetry. If, when our knowledge base is K, we choose the "closest" solution $\vec{x}$ to $\vec{X}$, then the agent's knowledge is having the least possible impact on their initial beliefs. However, although Euclidean distance is an obvious choice of metric there does not seem to be much ideological support behind it, which would favour it over other metrics.

We can rearrange the definition of MD^L, using $\sum_{i=1}^{J} x_i = 1$ (as in [ParVen2]), to show that

$$MD^L(K) = \text{ the unique } \vec{x} \in V^L(K) \text{ for which } \sum_{i=1}^{J} x_i^2 \text{ is minimal} \tag{1.75}$$

so we recognise MD^L as none other than Ren_2^L. From now on, we usually refer to MD^L as MD, since the *Renyi Processes* satisfies Language Invariance ([Moh]).

1.5.5 The *Centre of Mass* inference process and CM_∞

Definition ([ParVen2]) The *Centre of Mass inference process* on a language L, CM^L is defined by

$$CM^L(K) = \text{ the } \vec{x} \text{ which is the centre of mass of } V^L(K) \tag{1.76}$$

assuming uniform density so that

$$CM^L(K)(\alpha_i) = \frac{\int_{V^L(K)} x_i \, dV}{\int_{V^L(K)} dV} \tag{1.77}$$

where the integrals are taken over the relative dimension of $V^L(K)$ using the same basis.

Since integration respects linear combinations, $CM^L(K) \in V^L(K)$.

The justification for CM^L is that the most representative solution of K is chosen and the centre of gravity of $V^L(K)$ is the most fairly weighted estimate of the "true" probabilities. The assumption that there is some "true" p.b.f. which the rational agent is trying to estimate is a different approach to inference. The *principle of indifference* suggests that every point in $V^L(K)$ should be equally weighted.

The main criticism of CM^L is that it is not Language Invariant, so adding extra propositional variables to the overlying language affects the inferred probabilities of sentences of the original language.

Dirichlet priors can be used instead of the uniform distribution in order to get Language Invariance—see [LawWil]. However even in the case of the uniform distribution, if the number of propositional variables in the overlying language tends to infinity there is a limit which is a Language Invariant inference process.

Definition ([ParVen2]) The *limit centre of mass inference process for L*, denoted by CM^L_∞, is defined thus: $CM^L_\infty(K)$ is the unique $\vec{x} \in V^L(K)$ such that

$$\sum_{\substack{i=1, \\ i \notin I^L(K)}}^{J} \log x_i \tag{1.78}$$

is maximal. This is well defined since there exist solutions of K such as $ME(K)$ at which every co-ordinate not in $I^L(K)$ has a strictly positive value, because *Maximum Entropy* satisfies Open-mindedness.

Theorem 20 ([ParVen2]) *Let $K \in CL, \theta \in SL$. Then*

$$lim_{\substack{|L'| \to \infty \\ L \subseteq L'}} CM^{L'}(K)(\theta) = CM^L_\infty(K)(\theta) \tag{1.79}$$

The function (1.78) is concave i.e. the second derivative in each direction parallel to $\mathbb{D}^J$ from all points in $\{\vec{x} \in \mathbb{D}^J |\, x_i > 0 \text{ for all } i \notin I^L(K)\}$ is negative. However, K can vary such that $I^L(K)$ varies so we cannot use Theorem 13 to show that CM^L_∞ is continuous.

When we compare CM_∞^L with $Maximum\ Entropy$ using the Par-Ven Properties we obtain the following results, all of which are known to the authors of [ParVen2].

Theorem 21 CM_∞^L *satisfies Equivalence, Atomic Renaming, Obstinacy, Language Invariance, Open-mindedness and Relativisation. However, it fails to satisfy Continuity, Independence and Irrelevant Information.*

Proof Equivalence is satisfied because $CM_\infty^L(K)$ is defined in terms of $V^L(K)$.

Atomic Renaming is a property of CM_∞^L because the definition is symmetrical w.r.t. the atoms of L.

Obstinacy is satisfied because CM_∞^L chooses $\vec{x} \in V^L(K)$ to be maximal w.r.t. a fixed partial ordering, using Theorem 9. For that ordering, we first compare vectors by how many zeros they contain (fewer zeros give preferable vectors), then by which has the higher value of

$$\sum_{i=1, x_i>0}^{J} \log\, x_i \tag{1.80}$$

so this ordering is independent of K.

See [[Par], page 74] for a proof that CM_∞^L satisfies Language Invariance—from now on, we usually refer to CM_∞ without mentioning the overlying language.

For Open-mindedness, see the above definition.

For Relativisation, we can use a proof essentially similar to that of Theorem "7.8" in [[Par], page 100].

The following lemmas show that Independence, Continuity and Irrelevant Information are not satisfied by CM_∞.

Lemma 22 CM_∞ *does not satisfy Independence.*

Proof We let $L = \{p_1, p_2, p_3\}$.

$$K = \left\{ Bel(p_1) = 1, Bel(p_2|p_1) = \frac{1}{3}, Bel(p_3|p_1) = \frac{1}{3} \right\} \tag{1.81}$$

Then if CM_∞ satisfies Independence, $CM_\infty(K)(p_2 \wedge p_3|p_1) = 1/9$. When we enumerate the atoms of L $\alpha_1, \ldots \alpha_8$ in the standard ordering and let $x_i = Bel(\alpha_i)$ for each

$i = 1, \dots 8$, we obtain

$$V^L(K) = \left\{ \left(\tau, \frac{1}{3} - \tau, \frac{1}{3} - \tau, \frac{1}{3} + \tau, 0, 0, 0, 0 \right) \text{ s.t. } 0 \le \tau \le \frac{1}{3} \right\} \tag{1.82}$$

To calculate $CM_\infty(K)$, we need to choose τ such that $\log(\tau) + 2\log(\frac{1}{3} - \tau) + \log(\frac{1}{3} + \tau)$ is maximal. In other words, we need to maximise $f(\tau) = \tau(\frac{1}{3} - \tau)^2(\frac{1}{3} + \tau)$. If $X = 1/9$, the derivative of f at $\tau = X$ is $16/6561 \neq 0$. Hence

$$CM_\infty(K) \neq \left(\frac{1}{9}, \frac{2}{9}, \frac{2}{9}, \frac{4}{9}, 0, 0, 0, 0 \right) \tag{1.83}$$

so CM_∞ does not satisfy Independence and we have proved the lemma.

■

Lemma 23 *CM_∞ is not continuous.*

Proof Let $L = \{p_1, p_2\}$. For all $\epsilon \in [0, 1]$, we let

$$K_\epsilon = \{Bel(p_1 \wedge \neg p_2) = \epsilon Bel(p_1 \wedge p_2), Bel(\neg p_1 \wedge \neg p_2) = 0\} \tag{1.84}$$

We enumerate the atoms of L $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in the standard ordering and let $x_i = Bel(\alpha_i)$ as usual. If $Sol(\epsilon)$ denotes $V^L(K_\epsilon)$, we obtain

$$Sol(\epsilon) = \left\{ (\tau, \epsilon\tau, 1 - \tau(1 + \epsilon), 0) \text{ s.t. } 0 \le \tau \le \frac{1}{1 + \epsilon} \right\} \tag{1.85}$$

which is the line segment with endpoints $(0, 0, 1, 0)$ and $\left(\frac{1}{1+\epsilon}, \frac{\epsilon}{1+\epsilon}, 0, 0\right)$, which are both continuous functions of ϵ. Hence, by Lemma 11, $K_\epsilon \to K_0$ as $\epsilon \searrow 0$.

However, to calculate $CM_\infty(K_\epsilon)$ we need to maximise $\log(\tau) + \log(\epsilon\tau) + \log(1 - \tau(1 + \epsilon))$ as τ varies in $[0, 1/(1 + \epsilon)]$. Hence we need to maximise

$$f(\tau) = \tau^2(1 - \tau(1 + \epsilon)) \tag{1.86}$$

and, by doing the differentiation, we find that $f(\tau)$ is maximal at $\tau = \frac{2}{3(1+\epsilon)}$ when $\epsilon > 0$. Thus, for all $\epsilon > 0$,

$$CM_\infty(K_\epsilon) = \left(\frac{2}{3(1 + \epsilon)}, \frac{2\epsilon}{3(1 + \epsilon)}, \frac{1}{3}, 0 \right) \tag{1.87}$$

and $lim_{\epsilon \searrow 0} CM_\infty(K_\epsilon) = (\frac{2}{3}, 0, \frac{1}{3}, 0)$. However, we can use Atomic Renaming to show that $CM_\infty(K_0) = (\frac{1}{2}, 0, \frac{1}{2}, 0)$. Thus CM_∞ is not continuous and we have proved the lemma.

■

Proof of Theorem 21 continued Since Irrelevant Information is not satisfied by CM_∞, by [[Par], page 89], we have proved the theorem.

■

Chapter 2

Some limit theorems for the *Renyi Processes*

2.1 A property of the Blaschke topology

The intuitive property of the Blaschke topology which we describe in the two theorems below will be used to prove limit theorems for inference processes in this chapter and also later in the thesis.

Recall from the Introduction that the Blaschke metric on convex sets induces the Blaschke topology on the equivalence classes of CL. The following theorem states that if a knowledge base is changed by moving one of the constants at the right hand side of a constraint, the knowledge embodied is changing continuously.

Theorem 24 *If K is a consistent knowledge base in CL then for all $\lambda \in \mathbb{R}$, we define*

$$K^{(\lambda)} = K \cup \left\{ \sum_{i=1}^{J} v_i x_i = \lambda \right\} \tag{2.1}$$

where $v_1, \dots v_J$ are real constants. Let $\lambda = \lambda_0$ be fixed. In the Blaschke topology, if $K^{(\lambda_0+\delta)}$ is consistent for each δ such that $0 \leq \delta \leq \delta_0$, $K^{(\lambda_0+\delta)} \to K^{(\lambda_0)}$ as $\delta \searrow 0$.

Proof See Appendix 1.

Theorem 25 *Let K be a fixed consistent knowledge base in CL and define $i_1 \dots i_k$ s.t. $1 \leq i_1 < i_2 \dots < i_k \leq J$. For all $\vec{\lambda} \in \mathbb{R}^k$, let $K(\vec{\lambda})$ denote $K \cup \{x_{i_1} = \lambda_1, x_{i_2} = \lambda_2, \dots x_{i_k} = \lambda_k\}$. Then, if $\vec{\lambda}$ varies such that $K(\vec{\lambda})$ is consistent, $K(\vec{\lambda})$ is a continuous function of $\vec{\lambda}$.*

Proof See Appendix 1.

2.2 On the limit as $r \to \infty$ of the *Renyi Processes*

Recall from the Introduction that, for real $r > 1$ and every knowledge base $K \in CL$,

$$Ren_r(K) = \text{the unique } \vec{x} \in \vec{V}^L(K) \text{ for which } \sum_{i=1}^{J} x_i^r \text{ is minimal} \tag{2.2}$$

Also recall from the Introduction that we can justify using these inference processes, for integer values of r greater than 1, by assuming that a scientist is conducting an experiment which it is possible to repeat any number of times. If we minimise the belief that the same outcome occurs in the first r trials, we use Ren_r. When $r = 2$ this gives the MD inference process.

Once we go beyond the two trials that are required to get an inference process, it seems arbitrary to choose a particular number of times to do the experiment. We should like to know what is deduced when r is arbitrarily large, if that question has a sensible answer. In fact $lim_{r \to \infty} Ren_r$ does exist! In the above discussion r is an integer, i.e. the number of trials. However, in the following argument it does not matter if r is allowed to vary among integer or real values in $(1, \infty)$.

Notation To define $Minimax^L$ we define $\tilde{}: \mathbb{R}^J \to \mathbb{R}^J$ given by: $\tilde{\vec{x}} =$ the unique vector $\tilde{\vec{x}}$ which is a permutation of $\vec{x}$ for which $\tilde{x}_1 \geq \tilde{x}_2 \geq \dots \geq \tilde{x}_J$.

Define the *minimax* ordering on vectors of $\mathbb{R}^J$ by:

$\vec{x}$ is before $\vec{y}$ in the *minimax* ordering iff $\tilde{\vec{x}} < \tilde{\vec{y}}$ lexicographically. We shall say that $\vec{x}$ is *minimax*-better than $\vec{y}$ in that case.

Definition For a consistent knowledge base K,

$Minimax^L(K) =$ the unique $\vec{x} \in V^L(K)$ which is $minimax-$best. When $Minimax^L$ is being calculated the abbreviation $Mmx^L(*)$ shall be used.

Range of notation The terms that refer to $Minimax^L$ explicitly are defined here for use throughout this thesis, namely the *minimax* ordering, *minimax*-better, *minimax*-best.

Theorem 26 *For all* $K \in CL$, $Ren_r(K) \to Minimax^L(K)$ *as* $r \to \infty$.

Remark Recalling that Ren_r is the inference process we use when we maximise the measure of uncertainty H_r, it is already known (see Chapter 1 of [Kap]) that the limit of the Renyi entropy is a function of the maximum value:

$$H_r(\vec{x}) = -\frac{\log(\sum_{i=1}^{J} x_i^r)}{r-1} \to -\log(max(\vec{x})) \tag{2.3}$$

as $r \to \infty$.

We shall prove the theorem above after we show that $Minimax^L$ is well-defined.

Since $\tilde{x}_1$ is the maximum value of $\vec{x}$ $Mmx^L(K)$ has the least possible maximum of the solutions—hence the name. Hence $Mmx^L(K)$ is chosen such that $\widetilde{Mmx^L}(K)_1$ is minimised and, among the solutions with that maximum, $\widetilde{Mmx^L}(K)_2$ is minimised and so on. Among the solutions $\vec{x}$ with minimal maximum $\tilde{x}_2$ is minimised, which can be thought of as the second largest value but we must be careful: if the maximum value occurs twice, $\tilde{x}_2 = \tilde{x}_1$. Thus $\tilde{x}_2$ is the largest number in $\vec{x}$ once one occurrence of the maximum has been removed.

Notation From this point on in the entire thesis the phrase "highest k values" of a vector refers to the vector's values of $\tilde{}_1, \tilde{}_2 \ldots \tilde{}_k$. Also as $\tilde{x}_1$ is the maximum value of a co-ordinate of $\vec{x}$ this will be referred to as $max(\vec{x})$.

We now show that $Minimax^L$ is well-defined.

Definition In this section, we assume that K is some consistent knowledge base in CL. In this chapter, and elsewhere if the context refers to $Minimax^L$, $m_1, m_2 \ldots m_J$ are functions of $K \in CL$ given by:

$$m_1 = min\{max_{1 \leq i \leq J} x_i \mid \vec{x} \in V^L(K)\} \tag{2.4}$$

and then

$m_{k+1} = min\{max\ x_i \mid$ there exist $i_1, i_2, \ldots i_k$ distinct from i and from each other and s.t. $x_{i_1}, x_{i_2} \ldots x_{i_k}$ are equal to $m_1, m_2, \ldots m_k$ respectively where $\vec{x} \in V^L(K)\}$.

Then we can see that $Mmx^L(K)$ is a vector in $\vec{V_L}(K)$ which is a permutation of $(m_1, m_2, \ldots m_J)$. The required minima above exist (rather than being merely infimums) because at each stage the set of vectors being considered is closed and bounded. Hence by the Heine-Borel Theorem it is compact. Also the function that takes the maximum value of a vector,

$$max : V^L(K) \to \mathbb{R} \tag{2.5}$$

is continuous using the Euclidean topology so because continuous images of compact sets are compact the "set of maximums" has a minimum.

From now on we will assume the existence of strict minima/maxima of continuous functions on compact sets such as $V(K)$ etc..

Lemma 27 *For every $k = 1, 2, \ldots J$, $m_1 \geq m_2 \ldots \geq m_k$ and there exists $\vec{x} \in V^L(K)$ s.t. $\tilde{x}_1 = m_1 \ldots \tilde{x}_k = m_k$. In other words, there are distinct co-ordinates $i_1, \ldots i_k$ such that $x_{i_p} = m_p$ for each $p = 1, \ldots k$ and no other value in $\vec{x}$ exceeds m_k .*

Proof By induction on k.

Base Case $\underline{k = 1}$ In this case the lemma is trivial.

Inductive Step Assume (Inductive Hypothesis) that in the case $k = p$ the lemma is true. Since there can exist p distinct atoms with beliefs $m_1, \ldots m_p$ respectively, where every other value is less than or equal to $m_p \leq \ldots \leq m_2 \leq m_1$, the least possible maximum of the others is not greater than m_p. Then there exists solutions of K such that the values of $p+1$ distinct atoms are $m_1, \ldots m_{p+1}$ and all other values are not greater then m_{p+1}. Thus the sequence $m_1, \ldots m_{p+1}$ is non-increasing and we have proved the lemma.

■

Theorem 28 *There exists a bijection* $\sigma : \{1, 2, \ldots J\} \to \{1, 2, \ldots J\}$ *such that for each* $k = 1, 2, \ldots J$ *and all* $\vec{x} \in V^L(K)$ *s.t.* $\tilde{x}_1 = m_1, \ldots \tilde{x}_k = m_k$ *then* $x_{\sigma(i)} = m_i$ *for all* $i = 1, 2, \ldots k$.

Definition If, given K, the identity permutation can fulfil the role of σ above, we say that K *admits the identity permutation w.r.t. Theorem* 28.

As we learn that a solution of K has its largest beliefs in atoms equal to $m_1, m_2, \ldots$ we know that $x_{\sigma(1)} = m_1$, $x_{\sigma(2)} = m_2$ etc. The following claim is proved by induction on a.

Claim(a) There exist distinct $\sigma(1), \sigma(2) \ldots \sigma(a)$ in $\{1, \ldots J\}$, such that for every $k = 1, \ldots a$ and every solution $\vec{x}$ of K for which $\tilde{x_1} = m_1, \ldots \tilde{x_k} = m_k$, $x_{\sigma(i)} = m_i$ for each $i = 1, 2, \ldots k$.

Proof

Base Case $\underline{a = 1}$. We must show that if $\vec{x} \in V^L(K)$ such that $max(\vec{x}) = m_1$, there exists a certain $\sigma(1)$ such that $x_{\sigma(1)} = m_1$ for all such $\vec{x}$. Now suppose for contradiction that no such $\sigma(1)$ exists, then there exists a solution of K, $\vec{X}^{(1)}$, such that $max(\vec{X}^{(1)}) = m_1$ and $X_1^{(1)} < m_1$ and for the other $i = 2, 3, \ldots J$ there exists $\vec{X}^{(i)} \in V^L(K)$, s.t. $max(\vec{X}^{(i)}) = m_1$ and $X_i^{(i)} < m_1$.

Since $V^L(K)$ is convex,

$$\vec{y} = \frac{1}{J} \sum_{j=1}^{J} \vec{X}^{(j)} \tag{2.6}$$

is a solution of K.

Let $i \in \{1, \ldots J\}$. Then $Jy_i = \sum_{j=1}^{J} X_i^{(j)}$. Since every $\vec{X}^{(j)}$ has maximum m_1, every term in the sum is not greater than m_1. Also $X_i^{(i)} < m_1$ so $Jy_i < \sum_{i=1}^{J} m_1$. Hence $y_i < m_1$ for each $i = 1, \ldots J$ and the maximum of $\vec{y}$ is below the least possible maximum of solutions of K, so we have found a contradiction.

Hence such a $\sigma(1)$ can be fixed.

Inductive Step Suppose (I.H.) that Claim(t) is true so that $\sigma(1), \sigma(2), \ldots, \sigma(t)$ exist as above. We now show that $\sigma(t+1)$ exists between 1 and J, distinct from $\sigma(1) \ldots \sigma(t)$ such that if $\vec{x}$ is a solution of K with its highest $t+1$ values equal to

$m_1, \ldots m_{t+1}$ then $x_{\sigma(t+1)} = m_{t+1}$.

For all $\vec{x} \in V^L(K)$ for which $\tilde{x}_i = m_i$ for $i = 1, \ldots t$ we have by the I.H. that $x_{\sigma(1)} = m_1, \ldots x_{\sigma(t)} = m_t$.

The largest value at the other co-ordinates is m_{t+1}. If we suppose for contradiction that no fixed $\sigma(t+1)$ must equal m_{t+1} for all relevant $\vec{x}$, we take an arithmetic mean in a similar manner to the proof of the Base Case, giving a solution of K in which the top t values are $m_1, \ldots m_t$ but the next highest is smaller than m_{t+1}. This contradicts the definition of m_{t+1}, so the Inductive Step follows and we have proved the theorem.

■

Corollary 29 *The inference process $Minimax^L$ is well-defined.*

Proof Apply $k = J$ to the theorem, then there exists a permutation σ of $\{1, 2, \ldots J\}$ such that if $\vec{x} \in V^L(K)$ is a permutation of $(m_1, \ldots m_J)$ then $\vec{x} = (m_{\sigma^{-1}(1)}, m_{\sigma^{-1}(2)}, \ldots m_{\sigma^{-1}(J)})$ is forced. If $\vec{X} = Minimax^L(K)$ we know that $\vec{X} \in V^L(K)$ is a permutation of $(m_1, \ldots m_J)$ so the value of $\vec{X}$ is forced and we have proved the corollary.

■

Notation In general, let $\vec{x}^{(r)}$ denote $Ren_r(K)$ so that $x_i^{(r)} = Ren_r(K)(\alpha_i) = Ren_r(K)_i$ and $max^{(r)}$ will denote $max(\vec{x}^{(r)}) = max_{1 \leq i \leq J} x_i^{(r)}$.

Proof of Theorem 26 Let $K \in CL$ be fixed and we assume w.l.o.g. that K admits the identity permutation w.r.t. Theorem 28. We can do this because $Minimax^L$ satisfies Atomic Renaming and so do the Ren_r inference processes. Lemma 30 below states that the maximum value at the co-ordinates is tending to its minimum allowed value.

Lemma 30 *As $r \to \infty$, $max^{(r)} \to m_1$*

Proof By definition of m_1, $m_1 \leq max^{(r)}$ for every $r \in \mathbb{N}$, so it is sufficient to prove that for all positive $\epsilon < m_1$

$$\text{there exists } N \in \mathbb{N} \text{ such that for all } r > N, max^{(r)} < m_1 + \epsilon \tag{2.7}$$

Choose $N \geq \frac{\log J}{\log(\frac{1}{1-\epsilon})}$. Then

$$\begin{aligned} r > N \Rightarrow r \log\left(\frac{1}{1-\epsilon}\right) &> \log J && (2.8)\\ \Rightarrow \log\left(\frac{1}{1-\epsilon}\right) &> \frac{\log J}{r} && (2.9)\\ \Rightarrow \log(1-\epsilon) &< -\frac{\log J}{r} && (2.10)\\ \Rightarrow 1-\epsilon &< J^{-\frac{1}{r}} && (2.11) \end{aligned}$$

Hence, for all $x \in [m_1, 1]$,

$$\frac{x-\epsilon}{x} < J^{-\frac{1}{r}} \tag{2.12}$$

so raising both sides to the rth power implies that

$$\left(\frac{x-\epsilon}{x}\right)^r < J^{-1} \tag{2.13}$$

and so

$$(max^{(r)})^r > J(max^{(r)} - \epsilon)^r \tag{2.14}$$

by the definition of m_1. Hence, for every $\vec{y} \in V^L(K)$ for which $max(\vec{y}) \leq max^{(r)} - \epsilon$,

$$\sum_{i=1}^{J} (x_i^{(r)})^r > J(max^{(r)} - \epsilon)^r \geq \sum_{i=1}^{J} {y_i}^r \tag{2.15}$$

If such a $\vec{y}$ exists, this is a contradiction due to the definition of $Ren_r(K)$. Hence the maximum of a solution of K must be greater than $max^{(r)} - \epsilon$, completing the proof of Lemma 30.

■

To prove Theorem 26, we still need to show that the beliefs inferred for *individual* atoms tend to appropriate limits, not just $max(Ren_r(K))$.

Definition In this chapter only, for each integer $k = 1, \ldots J$, define $g_k : CL \to \mathbb{R}$ thus:

$$g_k = min\{max_{k \leq i \leq J} x_i | \, \vec{x} \in V^L(K)\} \tag{2.16}$$

Lemma 31 *The functions g_k (of which g_1 is also called m_1) are uniformly continuous functions in the Blaschke topology. In fact, for each $k = 1, 2, \ldots J$ and all $K_1, K_2 \in CL$, $|g_k(K_1) - g_k(K_2)| \leq \Delta(K_1, K_2)$.*

Proof We use the property of the function max that for all $\epsilon > 0$, $|\vec{x}-\vec{y}| < \epsilon$ implies that $|max(\vec{x}) - max(\vec{y})| < \epsilon$. Suppose that K_1, $K_2 \in CL$ and $\Delta(K_1, K_2) < \epsilon$. Let $Mmx^L(K_1) = \vec{k}$ and $Mmx^L(K_2) = \vec{p}$. Now there exists $\vec{x} \in V^L(K_1)$ s.t. $|\vec{x} - \vec{p}| < \epsilon$. Then $|max(\vec{x}) - m_1(K_2)| < \epsilon$. Similarly there exists $\vec{y} \in V^L(K_2)$ s.t. $|\vec{y} - \vec{k}| < \epsilon$ so that $|max(\vec{y}) - m_1(K_1)| < \epsilon$. Then

$$m_1(K_1) \leq max(\vec{x}) < m_1(K_2) + \epsilon$$

and

$$m_1(K_2) \leq max(\vec{y}) < m_1(K_1) + \epsilon \tag{2.17}$$

so $|m_1(K_1) - m_1(K_2)| < \epsilon$ and m_1 is *uniformly* continuous on CL. By letting $\epsilon \searrow \Delta(K_1, K_2)$, we deduce that

$$|g_k(K_1) - g_k(K_2)| \leq \Delta(K_1, K_2) \tag{2.18}$$

Similarly we can show that $\Delta(K_1, K_2) < \epsilon \Rightarrow |g_k(K_1) - g_k(K_2)| < \epsilon$ for each k s.t. $1 \leq k \leq J$ and all $K_1, K_2 \in CL$. We have proved the lemma.

■

Proof of Theorem 26 continued Now it is sufficient for us to show, by induction on i, that as $r \to \infty$, $x_i^{(r)} \to m_i$ for each $i = 1, 2, \ldots J$.

Base Case $\underline{i = 1}$ To show that $x_1^{(r)} \to m_1$ as $r \to \infty$ suppose not, for contradiction. Then for a fixed $\delta > 0$ and for every $N \in \mathbb{N}$ there exists $r(N) > N$ for which $|x_1^{(r)} - m_1| > \delta$. Consider the sequence $x_1^{(r(N))}$ for $N \geq 1$. By the compactness of $[0, 1]$, this sequence has a convergent subsequence $x_1^{(r(i))}$ with a limit $l \neq m_1$. Since $V^L(K)$ is closed and bounded it is compact (Heine-Borel Theorem) so the sequence $\vec{x}^{(r(i))}$ has a convergent subsequence, say $\vec{x}^{(r(j))}$ with limit $\vec{X}$. Now, by Lemma 30, $max(\vec{X}) = m_1$ since max is continuous, and $X_1 = l \neq m_1$. However $\vec{X} \in V^L(K)$ and $max(X) = m_1$ but $Mmx^L(K) = (m_1, m_2, \ldots m_J)$, and we have proved a contradiction. Hence $x_1^{(r)} \to m_1$ as required.

Inductive Step Suppose (I.H.) that $x_1^{(r)} \to m_1, \ldots x_k^{(n)} \to m_k$ as $r \to \infty$, for some k s.t. $1 \leq k < J$.

Case 1 $\underline{m_{k+1} = 0.}$ Then $m_{k+2} = \ldots = m_J = 0$ and as $r \to \infty$,

$$\sum_{i=k+1}^{J} x_i^{(r)} = \left(1 - \sum_{i=1}^{k} x_i^{(r)}\right) \to 1 - m_1 - m_2 \ldots - m_k$$
$$= m_{k+1} + m_{k+2} \ldots + m_J = 0 \tag{2.19}$$

Now $\sum_{i=k+1}^{J} x_i^{(r)} \geq x_{k+1}^{(r)} \geq 0$ so by the "Sandwich Rule" for limits $x_{k+1}^{(r)} \to 0$ as $r \to \infty$.

■

Case 2 $\underline{m_{k+1} > 0.}$ Let

$$K^{(r)} = K + \left\{x_1 = x_1^{(r)}, x_2 = x_2^{(r)} \ldots x_k = x_k^{(r)}\right\} \tag{2.20}$$

and

$$K^{(\infty)} = K + \{x_1 = m_1 \ldots x_k = m_k\} \tag{2.21}$$

Recall that, if $1 \leq k \leq J$, $g_k(K) = min\{max_{i=k+1\ldots J} x_i | \vec{x} \in V^L(K)\}$. By the Inductive Hypothesis and Theorem 25, $K^{(r)} \to K^{(\infty)}$ as $r \to \infty$. Given ϵ, a positive real number less than

$$F = \frac{(m_{k+1} + \ldots m_J)}{J - k} \tag{2.22}$$

there exists $N_1 \in \mathbb{N}$ such that for every $r > N_1$,

$$\Delta(K^{(r)}, K^{(\infty)}) < \epsilon/2 \implies |g_k(K^{(r)}) - g_k(K^{(\infty)})| < \epsilon/2 \tag{2.23}$$

(by Lemma 31) $\Rightarrow |g_k(K^{(r)}) - m_{k+1}| < \epsilon/2$. Also since

$$\lim_{r\to\infty} \sum_{i=k+1}^{J} x_i^{(r)} = 1 - \sum_{i=1}^{k} m_i = \sum_{i=k+1}^{J} m_i \tag{2.24}$$

there exists $N_2 > N_1 \in \mathbb{N}$ such that for every $r > N_2$

$$\sum_{i=k+1}^{J} x_i^{(r)} > (J - k)F/2 \tag{2.25}$$

so that

$$max^{(k,r)} = max_{i=k+1}^{J} x_i^{(r)} > F/2 > \epsilon/2 \tag{2.26}$$

Now suppose that $r > max\left(N_2, \frac{\log J}{\log(\frac{2}{2-\epsilon})}\right)$. In the same way as (2.8) this ensures that

$$1 - \epsilon/2 < J^{-\frac{1}{r}} \tag{2.27}$$

which implies that

$$\frac{x - \epsilon/2}{x} < J^{-\frac{1}{r}} \tag{2.28}$$

where $x = max^{(k,r)}$. Hence

$$(max^{(k,r)})^r > J(max^{(k,r)} - \epsilon/2)^r \tag{2.29}$$

so if a solution of K, say $\vec{y}$, exists for which $y_1 = x_1^{(r)} \dots y_k = x_k^{(r)}$ and its maximum value at the other co-ordinates is at least $\epsilon/2$ less than that of $Ren_r(K)$,

$$\sum_{i=1}^{J} y_i^r < \sum_{i=1}^{J} x_i^{(r)^r} \text{ and } Ren_r(K) = Ren_r(K^{(r)}) \tag{2.30}$$

by Obstinacy ([Moh]), which is a contradiction. Thus $|g_k(K^{(r)}) - max^{(r,k)}| < \epsilon/2$. Combining this with (2.23) gives

$$|max^{(k,r)} - m_{k+1}| < \epsilon \tag{2.31}$$

so $max^{(k,r)} \longrightarrow m_{k+1}$ as $r \longrightarrow \infty$.

Suppose now that $x_{k+1}^{(r)}$ does not tend to m_{k+1}. Then, similarly to the Base Case, there exists a subsequence of $x_{k+1}^{(r)}$ bounded away from m_{k+1} and by compactness a subsequence of that converges to some limit $l \neq m_{k+1}$. The corresponding sequence of $\vec{x}^{(r)}$ has a convergent subsequence with limit vector $\vec{X}$ for which $X_1 = m_1, \dots X_k = m_k$ by the I.H. and since $max_{i=k+1}^{J} x_i$ is continuous, $max_{i=k+1}^{J} X_i = m_{k+1}$ but $X_{k+1} = l$.

Since $\sigma(k+1) = k+1$ we deduce a contradiction. Hence $x_{k+1}^{(r)} \longrightarrow m_{k+1}$ as $r \longrightarrow \infty$, completing the Inductive Step and the proof of Theorem 26.

■

2.3 On the limit of the *Renyi Processes* as $r \to 0$

Recall from the Introduction that in [Moh] it was shown that the *Renyi Processes* form a continuum of continuous inference processes Ren_r for $r \in (0, \infty)$. We have seen in Section 2.1 that as $r \to \infty$ the limit is $Minimax^L$. The corresponding result at the opposite end of the continuum is:

Theorem 32 *For all $K \in CL$, $lim_{r \searrow 0} Ren_r(K) = CM_\infty(K)$.*

Notation Recall from the Introduction that, for all $r \in (0, 1)$ and each $K \in CL$,

$$Ren_r(K) = \text{the unique } \vec{x} \in V^L(K) \text{ for which } \sum_{i=1}^{J} x_i^r \text{ is maximal} \tag{2.32}$$

Notation For the rest of this section, we use a fixed knowledge base $K \in CL$ and express it in *dezeroed form*. This means that we label the atoms of L $\alpha_1, \ldots \alpha_J$ in such a way that $I^L(K) = \{i \text{ s.t. } J' < i \leq J\}$ where J' is some integer such that $2 \leq J' \leq J$.

As usual we let $x_i = Bel(\alpha_i)$ for each $i = 1, \ldots J$. Since the inference processes we use satisfy Atomic Renaming and Equivalence, (by [Moh], Theorem 21), we assume w.l.o.g. that K includes constraints of the form $x_i = 0$ for each i s.t. $J' < i \leq J$ and that every other constraint only refers to $x_1, x_2, \ldots x_{J'}$. We write solutions of K as points in $\mathbb{D}^{J'}$, ignoring the constant zeroes. Hence

$$Ren_r(K) = \text{the unique } \vec{x} \in V^L(K) \text{ for which } \sum_{i=1}^{J'} x_i^r \text{ is maximal} \tag{2.33}$$

for all $r \in (0, 1)$. Also, the definition (1.78) now gives us that

$$CM_\infty(K) = \text{the unique } \vec{x} \in V^L(K) \text{ for which } \sum_{i=1}^{J'} \log x_i \text{ is maximal} \tag{2.34}$$

since $\{1, \ldots J'\} = \{1 \ldots J\} \setminus I^L(K)$. With $CM_\infty(K)$, $Ren_r(K)$, written as vectors in $\mathbb{D}^{J'}$, the statement of Theorem 32 does not change. We need only show that

$$Ren_r(K)_i \to CM_\infty(K)_i \text{ for each } i = 1, \ldots J' \tag{2.35}$$

since convergence is trivial at those co-ordinates in $I^L(K)$.

It is now useful for us to show that the beliefs inferred for $\alpha_1, \ldots \alpha_{J'}$ by *Renyi Processes* as $r \searrow 0$ are bounded away from zero.

Lemma 33 *There exists $\delta > 0$ such that for every $r \in (0, \frac{1}{2})$ and each $i = 1, \ldots J'$, $Ren_r(K)_i \geq \delta$.*

Proof Let $\vec{x}^{(r)} = Ren_r(K) \in \mathbb{D}^{J'}$ for all $r \in (0, 1)$—we use this notation for the rest of Section 2.2. Also let $\vec{X} = CM_\infty(K) \in \mathbb{D}^{J'}$ and define $min_{i=1,\ldots J'} X_i = m$, which is positive since CM_∞ satisfies Open-mindedness, by Theorem 21. Now we let $\delta = \frac{m^3}{4}$ and suppose for contradiction that for some $i_1 = 1$ or $\ldots$ or J', and a fixed $R \in (0, \frac{1}{2})$, $x_{i_1}^{(R)} < \delta$. Let $\vec{x'}$ denote $\vec{x}^{(R)}$, so that $x'_{i_1} < \delta$.

Consider the line connecting $\vec{x'}$ and $\vec{X}$ given by

$$\{\vec{w}(\tau) = \vec{x'} + \tau(\vec{X} - \vec{x'}) \,|\, \tau \in \mathbb{R}\} \tag{2.36}$$

and the polynomial function of τ

$$F(\tau) = \sum_{i=1}^{J'} (w_i(\tau))^R \tag{2.37}$$

Differentiating F at $\tau = 0$ gives

$$F'(0) = R \sum_{i=1}^{J'} {x'_i}^{R-1} (X_i - x'_i) \tag{2.38}$$

We now show that $F'(0)$ is positive, to contradict the fact that $\vec{x'} = Ren_R(K)$.

Since $m < 1$,

$$x'_{i_1} < \frac{m^3}{4} < \frac{m}{2} < m \leq X_{i_1} \tag{2.39}$$

so the term of the sum (2.38) with $i = i_1$ is positive. In general a term of the sum (2.38) with $i = i*$ is negative iff $x'_{i*} > X_{i*}$ so let the $i* \in \{1, \ldots J'\}$ for which this holds be enumerated by $q_1, \ldots q_s$.

For each $j = 1, \ldots s$, $X_{q_j} \geq m$ so $x'_{q_j} > m$ and $x'_{i_1} < \delta = \frac{m^3}{4}$ so $\frac{x'_{q_j}}{x'_{i_1}} > m/(\frac{m^3}{4})$. Simplifying this gives

$$\frac{x'_{q_j}}{x'_{i_1}} > \left(\frac{2}{m}\right)^2 \tag{2.40}$$

Using $R - 1 < -\frac{1}{2}$,

$$\left(\frac{x'_{q_j}}{x'_{i_1}}\right)^{R-1} < \frac{m}{2} \tag{2.41}$$

Also $\delta < \frac{m}{2}$, so $X_{i_1} - x'_{i_1} > m - \delta > \frac{m}{2}$. Now

$$\sum_{j=1}^{s}(x'_{q_j} - X_{q_j}) < \sum_{j=1}^{s} x'_{q_j} < 1 \tag{2.42}$$

so

$$\frac{m}{2}\sum_{j=1}^{s}\frac{x'_{q_j} - X_{q_j}}{X_{i_1} - x'_{i_1}} \;<\; \frac{m}{2}\sum_{j=1}^{s}\left(\frac{2}{m}\right)(x'_{q_j} - X_{q_j}) \;<\; 1 \tag{2.43}$$

Applying (2.41) gives

$$\sum_{j=1}^{s}\left(\frac{x'^{R-1}_{q_j}(x'_{q_j} - X_{q_j})}{x'^{R-1}_{i_1}(X_{i_1} - x'_{i_1})}\right) < 1$$

$$\Rightarrow x'^{R-1}_{i_1}(X_{i_1} - x'_{i_1}) > \sum_{j=1}^{s} x'^{R-1}_{q_j}(x'_{q_j} - X_{q_j}) \tag{2.44}$$

When we multiply this by R, the left hand side is a positive term in the sum (2.38) and the right hand side is the total positive modulus of all the negative terms.

Thus $F'(0) > 0$ so there exists $\epsilon_0 \in (0,1)$ such that $F(\epsilon) > F(0)$ for every $\epsilon \in (0, \epsilon_0)$. Let

$$\vec{z} = \vec{x'} + \frac{\epsilon}{2}(\vec{X} - \vec{x'}), \text{ then } \sum_{i=1}^{J'} z_i^R > \sum_{i=1}^{J'} x'^R_i \tag{2.45}$$

Now $\vec{z} \in V(K)$ by convexity since $\vec{x'}$ and $\vec{X}$ are in in $V(K)$ so this contradicts the fact that $\vec{x'} = Ren_R(K)$. Hence we have proved the lemma.

■

Proof of Theorem 32 continued

Definition For all $r \in (0, \frac{1}{2})$ we let $G_r, G_0 : [\delta, 1]^{J'} \to \mathbb{R}$ be given by:

$$G_r(\vec{x}) = \sum_{i=1}^{J'}\frac{x_i^r - 1}{r},\; G_0(\vec{x}) = \sum_{i=1}^{J'}\log_e(x_i) \tag{2.46}$$

By definition of Ren_r, $Ren_r(K)$ = the unique $\vec{x} \in V(K)$ for which $G_r(\vec{x})$ is maximal. Also $CM_\infty(K)$ = the unique $\vec{x} \in V(K)$ for which $G_0(\vec{x})$ is maximal. Intuitively we can think of the following lemma as useful in a proof of Theorem 32.

Lemma 34 *As* $r \searrow 0$ $G_r - G_0 \to 0$ *uniformly.*

Proof of lemma Firstly we show that $g_r(x) = \frac{x^r-1}{r}$ uniformly converges to $g_0(x) = \log_e(x)$ as $r \searrow 0$ for $x \in [\delta, 1]$. Indeed

$$g_r(x) = -\int_x^1 v^{r-1}dv \quad \text{and} \quad g_0(x) = -\int_x^1 v^{-1}dv \tag{2.47}$$

For all v such that $\delta \leq v \leq 1$,

$$|v^{r-1} - v^{-1}| = |v^{-1}(v^r - 1)| \leq \frac{1-\delta^r}{\delta} \tag{2.48}$$

which tends to 0 as $r \searrow 0$ independently of v. Since the range of integration above is bounded by 1, the modulus of the difference between the integrals is bounded above by $\frac{1-\delta^r}{\delta}$ as well. Hence g_r uniformly converges to g_0. Taking sums across J' co-ordinates each in $[\delta, 1]$ preserves uniform convergence so $G_r(\vec{x}) - G_0(\vec{x}) \to 0$ uniformly as $r \searrow 0$.

■

Proof of Theorem 32 continued Recall that $\vec{x}^{(r)}$ denotes $Ren_r(K)$ for all $r \in (0, \frac{1}{2})$ and $\vec{X}$ denotes $CM_\infty(K)$. We need to prove that $\vec{x}^{(r)} \to \vec{X}$ as $r \searrow 0$ to show Theorem 32, by (2.35). Assume for contradiction that $\vec{x}^{(r)}$ doesn't tend to $\vec{X}$ as $r \searrow 0$. Then there exists $\xi > 0$ such that for some arbitrarily small positive values of r, $|\vec{x}^{(r)} - \vec{X}| > \xi$. There must be a convergent subsequence of these $\vec{x}^{(r)}$, indexed by $r = r_p$ because $\mathbb{D}^{J'}$ is compact so

$$\lim_{r_p \searrow 0} (\vec{x}^{(r_p)}) = \vec{b} \neq \vec{X} \tag{2.49}$$

Since the $\vec{x}^{(r_p)}$ are in $E = V(K) \cap [\delta, 1]^{J'}$, a closed set, $\vec{b} \in E$. Recall that $\vec{X}$ has every co-ordinate value greater than δ. Choose $\eta > 0$ such that

$$G_0(\vec{X}) - G_0(\vec{b}) > \eta \tag{2.50}$$

Now, by Lemma 34, we fix $\rho > 0$ such that for every $r < \rho$ and all $\vec{x} \in E$,

$$|G_0(\vec{x}) - G_r(\vec{x})| < \frac{\eta}{4} \tag{2.51}$$

If r_p satisfies $r_p < \rho$ and $|G_0(\vec{x}^{(r_p)}) - G_0(\vec{b})| < \eta/4$ then

$$G_{r_p}(\vec{X}) \;\; \geq \;\; G_0(\vec{X}) - \frac{\eta}{4}$$

$$
\begin{aligned}
> G_0(\vec{b}) + \frac{3\eta}{4} &\geq G_0(\vec{x}^{(r_p)}) - \frac{\eta}{4} + \frac{3\eta}{4} \\
&\geq G_{r_p}(\vec{x}^{(r_p)}) - \frac{\eta}{4} - \frac{\eta}{4} + \frac{3\eta}{4} > G_{r_p}(\vec{X}) \qquad (2.52)
\end{aligned}
$$

so we have proved a contradiction. Hence $lim_{r\searrow 0}(Ren_r(K)) = CM_\infty(K)$ and we have proved the theorem.

■

Chapter 3

The properties of $Minimax^L$

3.1 Comparing $Minimax^L$ with $Maximum\ Entropy$

In this section, we test $Minimax^L$ against the Par-Ven Properties defined in Subsections 1.4.1-1.4.9.

3.1.1 Equivalence

Theorem 35 *$Minimax^L$ satisfies Equivalence.*

Proof This property holds for $Minimax^L$ since the inference process is defined in terms of $V^L(K)$.

■

3.1.2 Atomic Renaming

Theorem 36 *$Minimax^L$ satisfies Atomic Renaming.*

Proof The definition of $Minimax^L$ is symmetrical with respect to permutations of the atoms, so Atomic Renaming holds for $Minimax^L$.

■

3.1.3 Obstinacy

Theorem 37 *$Minimax^L$ satisfies Obstinacy.*

Proof For any consistent knowledge base K, $Mmx^L(K)$ is the optimal solution of K w.r.t. a fixed partial ordering, namely the *minimax* ordering. Hence, by Theorem 9, $Minimax^L$ satisfies Obstinacy and we have proved the theorem.

■

Recall from the Introduction that an inference process is Obstinate if we can define it in the manner of (1.42). To express $Minimax^L$ in the form

$$\text{for all } K \in CL, Mmx^L(K) = \text{the unique } \vec{x} \in V^L(K) \text{ such that } F(\vec{x}) \text{ is minimal} \tag{3.1}$$

where $F : \mathbb{D}^J \to Q$ and Q is a totally ordered set, we can use the natural ordering on $Q = \mathbb{R}[\epsilon]$ where ϵ is an infinitesimal.

However, whereas most known Obstinate inference processes can be characterised by minimising a function taking values in $\mathbb{R}$, the next theorem shows that $Minimax^L$ cannot be defined in this manner.

Theorem 38 *When $|L| > 1$, no $F : \mathbb{D}^J \to \mathbb{R}$ exists such that $Minimax^L$ can be defined in the manner of* (3.1).

Proof For all real p, q such that $\frac{1}{2} \leq p \leq \frac{3}{5}$ and $q > 0$, define

$$K(p,q) = \{x_1 = p + qx_2,\, x_3 = 0,\, x_4 = 1 - x_1 - x_2\} \cup \{\, x_k = 0 \text{ s.t. } k \in \mathbb{N}, k > 4\} \tag{3.2}$$

When we write solution vectors, only the first 4 co-ordinates will be listed, the sum of the values of which must be 1. We say that $x_i = Bel(\alpha_i)$ as usual, where the atoms of L are enumerated in the standard ordering. Hence, if $|L| > 1$, only 2 propositional variables have variable beliefs and the others all have belief 1 at every solution.

We denote $V^L(K(p,q))$ by $Ans(p,q)$. Then

$$Ans(p,q) = \left\{ (p + q\tau, \tau, 0, (1-p) - \tau(1+q)) \,|\, 0 \leq \tau \leq \left(\frac{1-p}{1+q} \right) \right\} \tag{3.3}$$

Abbreviate $Mmx^L(K(p,q))$ to $Mmx(p,q)$ and assume for contradiction that a function $F = F_0$ has been fixed defining $Minimax^L$ as above. Then if $\vec{x} = Mmx(p,q)$ and $\vec{x} \neq \vec{y} \in Ans(p,q), F_0(\vec{x}) < F_0(\vec{y})$.

For all $p \in [\frac{1}{2}, \frac{3}{5}]$, define

$$upper(p) = F_0((p,0,0,1-p)),\ lower(p) = F_0\left(\left(p, \frac{1-p}{2}, 0, \frac{1-p}{2}\right)\right) \tag{3.4}$$

Only co-ordinates 2 and 4 do not take constant values in $Ans(p,0)$ and $x_2 + x_4$ is constant so, at $Mmx(p,0)$, $x_2 = x_4$ by Atomic Renaming and

$$Mmx(p,0) = \left(p, \frac{1-p}{2}, 0, \frac{1-p}{2}\right) = lower(p) \tag{3.5}$$

Also $upper(p) \in Ans(p,0)$, so $F_0(lower(p)) < F_0(upper(p))$ for every $p \in [\frac{1}{2}, \frac{3}{5}]$.

If $q > 0$ and $\tau > 0$, then $p + q\tau > p$ so the maximum of each solution of $K(p,q)$ is greater than p. Hence $Mmx(p,q) = (p,0,0,1-p)$ for all $q > 0$ and all $p \in [1/2, 3/5]$. We fix $p = p_0$ and let δ be positive and not greater than $\frac{3}{5} - p_0$. Then since

$$\left(p_0 + \delta, \frac{1-(p_0+\delta)}{2}, 0, \frac{1-(p_0+\delta)}{2}\right) \in Ans\left(p_0, \frac{2\delta}{1-(p_0+\delta)}\right), \tag{3.6}$$

$F_0(upper(p_0)) < F_0(lower(p_0+\delta))$ for all δ s.t. $p_0 < p_0 + \delta \leq \frac{3}{5}$.

Now a contradiction follows: for all $p \in [1/2, 3/5]$ choose $rat(p) \in \mathbb{Q} \cap (F_0(lower(p)), F_0(upper(p)))$, which exists by the density of $\mathbb{Q}$ in $\mathbb{R}$. All of these rational numbers $rat(p)$ must be different for different values of p since when $p < p'$, $F_0(upper(p)) < F_0(lower(p'))$. Hence we have found an uncountably infinite set of rational numbers, which is a contradiction, so no F can exist defining $Minimax^L$ in the above way and we have proved the theorem.

■

3.1.4 Language Invariance

Theorem 39 *$Minimax^L$ is Language Invariant.*

Proof Since every *Renyi Process* is Language Invariant (by [Moh]), then

$Mmx^L = lim_{r\to\infty}Ren_r^L$ (by Theorem 26) is also independent of L and we have proved the theorem. ■

From now on we will usually refer to the *Minimax* inference process without mentioning the overlying language.

3.1.5 Continuity

In this subsection we use m_1, m_2 etc. as defined in Chapter 2—recall (2.4). Since *Minimax* is a limit of continuous inference processes, it is certainly tempting to believe that it is continuous. Recall that, by Lemma 31, $m_1(K)$ is a uniformly continuous function in the Blaschke topology—a fact which can make continuity of *Minimax* seem very reasonable.

Also when a knowledge base varies continuously in certain ways *Minimax* varies continuously—we show this in Section 3.2. However, the function $m_2 : CL \to \mathbb{R}$ can be seen not to be continuous!

Theorem 40 *Minimax is not continuous.*

Proof For all $\epsilon \in [0, 1]$, we let

$$K_\epsilon = \left\{x_1 = \epsilon x_2,\ x_2 + x_3 = \frac{1}{2},\ x_1 + x_2 + x_3 + x_4 = 1\right\} \tag{3.7}$$

where $L = \{p_1, p_2\}$, $x_i = Bel(\alpha_i)$ for $i = 1, 2, 3, 4$ and the atoms α_i are labelled in the standard ordering. We let $\vec{s}(\epsilon, \tau)$ denote $(\epsilon\tau, \tau, \frac{1}{2} - \tau, \frac{1}{2} - \epsilon\tau)$. Let

$$Sol_\epsilon = V^L(K_\epsilon) = \left\{\vec{s}(\epsilon, \tau) \text{ s.t. } 0 \le \tau \le \frac{1}{2}\right\} \tag{3.8}$$

which is the line segment connecting $\vec{s}(\epsilon, 0) = (0, 0, \frac{1}{2}, \frac{1}{2})$ and $\vec{s}(\epsilon, \frac{1}{2}) = (\frac{\epsilon}{2}, \frac{1}{2}, 0, \frac{1-\epsilon}{2})$. The points $\vec{s}(\epsilon, 0)$ and $\vec{s}(\epsilon, \frac{1}{2})$ are continuous functions of ϵ. Hence by Lemma 11, $Sol_\epsilon \to Sol_0$ as $\epsilon \searrow 0$. Therefore $K_\epsilon \to K_0$ as $\epsilon \searrow 0$.

For $\epsilon > 0$, $max(\vec{s}(\epsilon, \tau)) =$ either τ or $\frac{1}{2} - \epsilon\tau$. With ϵ fixed, the maximum of these two values is minimal when they are equal, since $\frac{1}{2} - \epsilon\tau$ decreases as τ increases.

Hence for the *minimax*-best solution of K_ϵ, we require that $\tau = \frac{1}{2(1+\epsilon)}$. Thus

$$Mmx(K_\epsilon) = \left(\frac{\epsilon}{2(1+\epsilon)}, \frac{1}{2(1+\epsilon)}, \frac{\epsilon}{2(1+\epsilon)}, \frac{1}{2(1+\epsilon)}\right) \tag{3.9}$$

which tends to $(0, \frac{1}{2}, 0, \frac{1}{2})$ as $\epsilon \searrow 0$.

However, for $\epsilon = 0$, $Mmx(K_0) = (0, \frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ so $Minimax$ is not continuous and we have proved the theorem.

■

3.1.6 Open-mindedness

Theorem 41 *Minimax does not satisfy Open-mindedness.*

Proof Let

$$K = \left\{x_1 = 0,\ x_3 - x_2 = \frac{1}{2},\ x_1 + x_2 + x_3 + x_4 = 1\right\} \tag{3.10}$$

where $L = \{p_1, p_2\}$, $x_i = Bel(\alpha_i)$ for $i = 1, 2, 3, 4$ and the atoms α_i are labelled in the standard ordering. Then, if $\vec{x} \in V^L(K)$, $max(\vec{x}) = x_3$ since $x_3 \geq \frac{1}{2}$. This is minimised when $x_3 = \frac{1}{2}$ so

$$Mmx^L(K) = \left(0, 0, \frac{1}{2}, \frac{1}{2}\right) \tag{3.11}$$

However $(0, \frac{1}{8}, \frac{5}{8}, \frac{2}{8}) \in V^L(K)$, so $x_2 = 0$ is not necessarily true when $\vec{x} \in V^L(K)$. Hence $Minimax$ does not satisfy Open-mindedness and we have proved the theorem.

■

The following technical lemma can make calculating $Minimax$ easier and is also useful in analysing its properties.

Lemma 42 *Let $C \subset \{1, \ldots J\}$ and let $\sim$ be an equivalence relation on $\{1, \ldots J\} \setminus C$ such that the equivalence classes are all of equal size. Let $i_1, \ldots i_q$ be representatives of the q distinct equivalence classes. For each $\vec{x} \in \mathbb{D}^J$ s.t. $x_i = x_j$ for all i, j s.t. $i \sim j$, let $Simp(\vec{x}) = (x_{i_1}, \ldots x_{i_q})$. Then if $\vec{y} \in \mathbb{R}^J$ is such that $y_c = x_c$ for all $c \in C$, comparing $\vec{x}$ and $\vec{y}$ in the minimax ordering is equivalent to comparing $Simp(\vec{x})$ and $Simp(\vec{y})$ in the minimax ordering.*

Proof Let $\vec{x}, \vec{y}$ be as above and we let A denote the size of the equivalence classes w.r.t. $\sim$. If $Simp(\vec{x})$ and $Simp(\vec{y})$ are permutations of each other, clearly $\vec{y}$ is a permutation of $\vec{x}$. Otherwise suppose w.l.o.g. that $\widetilde{Simp(\vec{x})}_i = \widetilde{Simp(\vec{y})}_i$ for each i s.t. $i \leq k$ and $\widetilde{Simp(\vec{x})}_{k+1} < \widetilde{Simp(\vec{y})}_{k+1}$ for some $k \geq 0$. Then if there are v of the co-ordinates in C whose values in $\vec{x}$ (or equivalently in $\vec{y}$) are not less than $\widetilde{Simp(\vec{y})}_{k+1}$ then

$$\tilde{x}_i = \tilde{y}_i \text{ for all } i \leq v + Ak \text{ but } \tilde{x}_{(v+Ak+1)} < \tilde{y}_{(v+Ak+1)}. \tag{3.12}$$

Hence the *minimax* comparison of $\vec{x}$ and $\vec{y}$ matches that of $Simp(\vec{x})$ and $Simp(\vec{y})$ as required and we have proved the lemma.

■

Corollary 43 *Suppose $\vec{x}, \vec{y}$ are vectors in $\mathbb{R}^J$ such that $x_i = y_i$ for all $i \in C$, for some $C \subset \{1, \dots J\}$. W.l.o.g. let $C = \{1, \dots k\}$. We can do this by the symmetry of the minimax ordering w.r.t. permuting the co-ordinates. Then comparing $\vec{x}$ and $\vec{y}$ in the minimax ordering is equivalent to comparing $(x_{k+1}, \dots x_J)$ and $(y_{k+1}, \dots y_J)$. If the same C is a subset of co-ordinates which are constant w.r.t. a knowledge base $K \in CL$, then $Minimax(K) =$ that $\vec{x} \in V^L(K)$ for which $(x_{k+1}, \dots x_J)$ is minimax-best.*

Proof The first part of the corollary follows by using Lemma 42, letting $C = \{1, \dots, k\}$ and letting the relation $\sim$ on $\{k+1, \dots J\}$ be equality. For the second part, if there exists $\vec{y}$ *minimax*-better than $\vec{x}$, $(y_{k+1}, \dots y_J)$ is *minimax*-better than $(x_{k+1}, \dots x_J)$. We have proved the corollary.

■

3.1.7 Independence

Theorem 44 *$Minimax$ does not satisfy Independence.*

We contain the proof of Theorem 44 in a more general investigation of how "close" $Minimax$ is to satisfying Independence.

How close does $Minimax$ come to giving the Independent solution?

We will look later at the knowledge bases used in the definition of Independence.

For simplicity consider first a knowledge base $K_{b,c}$ of the form:

$$K_{b,c} = \{x_1 + x_2 = b,\ x_1 + x_3 = c\} \tag{3.13}$$

where $x_i = Bel(\alpha_i)$ for $i = 1, 2, 3, 4$.

Notation Let the *Independent solution* of $K_{b,c}$ be denoted by

$$Ind(b, c) = ME(K_{b,c}) = (bc, b(1-c), (1-b)c, (1-b)(1-c)) \tag{3.14}$$

for all $b \in [0, 1]$ and $c \in [0, 1]$. Also we denote $Minimax(K_{b,c})$ by $Mmx(b, c)$.

We can see that $K_{b,c}$ is consistent $\Leftrightarrow b \in [0, 1], c \in [0, 1]$. That is because the sum of two beliefs for different atoms must lie in $[0, 1]$ and if $b, c \in [0, 1]$ we can see that $ME(K_{b,c}) = Ind(b, c)$ is a solution. It is helpful, to avoid mentioning $max(b, c)$ etc., to establish the convention that

$$0 \leq b \leq c \leq \frac{1}{2} \tag{3.15}$$

We can do this without loss of generality because if b is greater than $\frac{1}{2}$, the constraint $x_1 + x_2 = b$ is equivalent to $x_3 + x_4 = 1 - b$ and this rearrangement will give $K_{1-b,c}$ up to the renaming of atoms that swaps x_3 for x_1 and x_2 for x_4. Also $1 - b < \frac{1}{2}$.

Similarly for the case $c > \frac{1}{2}$. Finally $b \leq c$ can be assumed, otherwise swapping them and exchanging the atoms α_2 and α_3 will produce an essentially similar $K_{b,c}$ in the form of (3.15). None of these changes will affect the distance from $Mmx(b, c)$ to $Ind(b, c)$.

Notation $Sol(b, c)$ shall denote $V^L(K_{b,c})$ in this subsection.

$$Sol(b, c) = \{(\tau, b - \tau, c - \tau, \tau + 1 - b - c) \mid 0 \leq \tau \leq b\} \tag{3.16}$$

That is because $1 - b - c \geq 0$ so the condition that $0 \leq \tau \leq b$, clearly required to make the first two co-ordinate values non-negative, also guarantees non-negativity at the other co-ordinates. The maximum of τ, $b - \tau, c - \tau$ and $\tau + 1 - b - c$ is

$max(c-\tau, \tau+1-b-c)$. Let $f(b,c) = \frac{1}{2}(2c+b-1)$, which is the value of τ for which $c-\tau = \tau+1-b-c$.

If $f(b,c) \in [0,b]$ then

$$Mmx(b,c) = (f, b-f, c-f, f+1-b-c) = (f, b-f, c-f, c-f) \tag{3.17}$$

because if $\tau < f$, $c-\tau$ will be larger (and maximal) but if $\tau > f$, $\tau+1-b-c$ will exceed $c-f$. If $f < 0$, $\tau+1-b-c$ is the maximum of every vector in $Sol(b,c)$ so $Mmx(b,c) = (0,b,c,1-b-c)$. Note that $f \leq b$ since $c \leq \frac{1}{2}$. In summary

$$\begin{aligned} \underline{Case\,1} : c \leq \frac{1-b}{2} \quad &\Rightarrow \quad Mmx(b,c) = (0,b,c,1-b-c) \\ \underline{Case\,2} : \frac{1-b}{2} \leq c \leq \frac{1}{2} \quad &\Rightarrow \end{aligned} \tag{3.18}$$

$$Mmx(b,c) = \left(\frac{2c+b-1}{2}, \frac{b-2c+1}{2}, \frac{1-b}{2}, \frac{1-b}{2}\right) \tag{3.19}$$

Theorem 45 *For $b \in [0,1]$, $c \in [0,1]$, $Mmx(b,c) = Ind(b,c)$ iff either b or c equal either 0, $\frac{1}{2}$ or 1.*

Proof We begin by assuming the convention, as before, that $0 \leq b \leq c \leq \frac{1}{2}$.

In Case 1, $bc = 0$ is necessary and since $b \leq c$, $bc = 0$ iff $b = 0$, $c \leq \frac{1-b}{2}$.

In Case 2, it is required that $\frac{1-b}{2} = (1-c)(1-b)$. In Case 2 $b \neq 1$ anyway so, upon checking that $c = \frac{1}{2}$ really does give $Ind = Mmx$ the solutions in Case 2 are given by $c = \frac{1}{2}$. Given the convention $0 \leq b \leq c \leq \frac{1}{2}$, $Ind(b,c) = Mmx(b,c) \Leftrightarrow b = 0$ or $c = \frac{1}{2}$.

When we allow b, c to take values throughout $[0,1]$, we deduce that $Ind(b,c) = Mmx(b,c)$ iff either b or c equal either 0, $\frac{1}{2}$ or 1 so we have proved the theorem.

■

Theorem 46 *If $b \in [0,1]$, $c \in [0,1]$, $|Mmx(b,c) - Ind(b,c)|$ takes its maximal value, $\frac{2}{9}$, iff $(b,c) = \left(\frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{2}{3}\right), \left(\frac{2}{3}, \frac{1}{3}\right)$ or $\left(\frac{2}{3}, \frac{2}{3}\right)$.*

Remark In terms of Euclidean distance, this is in fact the answer to the question: "In the worst scenarios, how far away is *Minimax* from satisfying Independence?" We answer this question later in the case of the knowledge bases used in the definition of Independence, but for now we use the $K_{b,c}$'s.

Proof Recall the notation of (3.16). Note that when we calculate distance between points in $Sol(b,c)$ we obtain:

$$\begin{aligned} &|(\tau, b-\tau, c-\tau, \tau+1-b-c) - (\upsilon, b-\upsilon, c-\upsilon, \upsilon+1-b-c)| \\ &= \sqrt{(\tau-\upsilon)^2 + (\tau-\upsilon)^2 + (\tau-\upsilon)^2 + (\tau-\upsilon)^2} = 2|\tau-\upsilon| \end{aligned} \tag{3.20}$$

Hence to maximise the distance just maximise the difference between the values at the first co-ordinate.

Now we again assume that $0 \leq b \leq c \leq \frac{1}{2}$, splitting cases as in (3.18).

<u>Case 1</u>: $\underline{c \leq \frac{1-b}{2}}$ The first co-ordinates of $Mmx(b,c)$ and $Ind(b,c)$ differ by bc. Let the triangular region R_1 be given by

$$R_1 = \left\{ (b,c) \text{ s.t. } 0 \leq b \leq c \leq \frac{1}{2} \text{ and } c \leq \frac{1-b}{2} \right\} \tag{3.21}$$

If, from some point of R_1 not on the edge $c = \frac{1}{2}(1-b)$, we move parallel to the $c-$axis, then we increase the c co-ordinate and increase bc. On the edge given by $c = \frac{1}{2}(1-b)$, $0 \leq b \leq \frac{1}{3}$ and so $\frac{d}{db}(bc) = \frac{1}{2} - b > 0$. Hence we see that for Case 1, $b = c = \frac{1}{3}$ is the "worst" point.

<u>Case 2</u>: $\underline{\frac{1-b}{2} \leq c \leq \frac{1}{2}}$ The first co-ordinates of $Mmx(b,c)$ and $Ind(b,c)$ differ by $|\frac{1}{2}(2c+b-1) - bc|$. Now for general real numbers b, c,

$$\frac{1}{2}(2c+b-1) = bc \iff b = 1 \text{ or } c = \frac{1}{2} \tag{3.22}$$

Since the function g, given by

$$g(b,c) = 2bc - (2c+b-1) \tag{3.23}$$

is continuous, the sign of that function cannot change in the region in which Case 2 applies, say R_2, given by

$$R_2 = \left\{ (b,c) \text{ s.t. } 0 \leq b \leq c \leq \frac{1}{2} \text{ and } \frac{1}{2} \geq c \geq \frac{1-b}{2} \right\} \tag{3.24}$$

Since g is positive at $(0, 0)$, it is positive or zero throughout R_2 so the problem is to maximise $g(b, c)$ in the triangle R_2.

It is clear that if b is fixed, $g(b, c)$ is a monotonic decreasing function of c. Also if c is fixed, g decreases monotonically when b varies. The search for points maximising g is now limited to the edge of R_2 given by

$$c = \frac{1-b}{2},\ 0 \leq b \leq \frac{1}{3}. \tag{3.25}$$

Now this is the same as the top edge of R_1 considered already, so $(\frac{1}{3}, \frac{1}{3})$ is the unique worst point for Case 2 as well.

Discarding the assumption that $0 \leq b \leq c \leq 1/2$, i.e. as we let b, c take values throughout $[0, 1]$, by the earlier discussion the knowledge bases $K_{b,c}$ producing the largest distances between $Ind(b, c)$ and $Mmx(b, c)$ are given by
$(b, c) = (\frac{1}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}), \frac{2}{3}, \frac{1}{3})$ or $(\frac{2}{3}, \frac{2}{3})$. The largest distance $|Ind(b, c) - Mmx(b, c)| = \frac{2}{9}$ is realised at just those points. We have proved the theorem.

■

Remark The worst case K_{bc}'s provide examples of $Minimax$ behaving very differently from Independence. For example, if $b = c = 1/3$, $Mmx(K_{bc})_1 = 0$.

Remark If we consider the overlying language to be variable, does this affect the relative distances between $Ind(b, c)$ and $Mmx(b, c)$?

The answer is an emphatic "No"! Just for this remark, we let $Ind(b, c)$ denote the p.b. function inferred from $K_{b,c}$ by an inference process satisfying Independence, Language Invariance and Atomic Renaming. Recall (Theorem 39) that $Minimax$ is Language Invariant. We write $Ind^L(b, c)$, $Mmx^L(b, c)$ for the vectors representing $Ind(b, c)$, $Mmx(b, c)$ respectively when $x_i = Bel(\alpha_i)$ and $\alpha_1, \ldots \alpha_J$ enumerates the atoms of L in the standard ordering.

Let $d^L(b, c)$ denote $|Ind^L(b, c) - Mmx^L(b, c)|$. Now, if $Mmx(b, c)^L = (X_1, \ldots X_J)$,

$$Mmx(b, c)^{L'} = \left(\frac{X_1}{2}, \frac{X_1}{2}, \ldots, \frac{X_J}{2}, \frac{X_J}{2}\right) \tag{3.26}$$

where $L' = L + \{p\}$. This is because the atoms $\alpha_1, \ldots \alpha_J$ of L have been replaced by atoms $\beta_1, \ldots \beta_{2J}$ where

$$\beta_{2l-1} = \alpha_l \wedge p,\ \beta_{2l} = \alpha_l \wedge \neg p \text{ for each } l = 1, \ldots J \tag{3.27}$$

For each $l = 1, \ldots J$, $Mmx(b,c)(\beta_{2l-1}) = Mmx(b,c)(\beta_{2l})$ by Atomic Renaming.

Similarly, if $Ind(b,c)^L = (Y_1, Y_2, \ldots Y_J)$, $Ind(b,c)^{L'} = \left(\frac{Y_1}{2}, \frac{Y_1}{2}, \ldots \frac{Y_J}{2}\right)$.

Then

$$\begin{aligned} (d^{L'}(b,c))^2 &= 2\sum_{i=1}^{J}\left(\frac{X_i - Y_i}{2}\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{J}(X_i - Y_i)^2 \\ &= \frac{1}{2}.(d^{L'}(b,c))^2 \end{aligned} \tag{3.28}$$

Adding a propositional variable to L has multiplied $d(b,c)$ by $\frac{1}{\sqrt{2}}$. Hence by induction on the number of extra propositional variables, the choices of (b,c) which maximise $|\,Ind(b,c) - Mmx(b,c)|$ are independent of the overlying language.

The knowledge bases used in the definition of Independence

Now, for all $a, b, c \in [0,1]$ s.t. $a > 0$ we define

$$K_{a,b,c} = \{Bel(p_1) = a, Bel(p_2|p_1) = b, Bel(p_3|p_1) = c\} \tag{3.29}$$

which is the knowledge base, (over the language $L = \{p_1, p_2, p_3\}$) given in the definition of Independence [[Par], page 101].

Notation The Independent solution, denoted $Ind(a,b,c)$, is the p.b. function inferred on $K_{a,b,c}$ by an inference process satisfying Independence and Atomic Renaming, i.e. $Ind(a,b,c) = ME(K_{a,b,c})$. In the following theorem let $Mmx(a,b,c)$ denote $Minimax(K_{a,b,c})$.

Theorem 47 *If $K_{a,b,c}$ is consistent, $|Mmx(a,b,c) - Ind(a,b,c)|$ takes its maximal value, $\frac{2}{9}$, when $a = 1$ and $(b,c) = \left(\frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{2}{3}\right), \left(\frac{2}{3}, \frac{1}{3}\right)$ or $\left(\frac{2}{3}, \frac{2}{3}\right)$.*

Proof We use the standard ordering of the atoms of L, letting $x_i = Bel(\alpha_i)$ as usual. The constraints of $K_{a,b,c}$ are

$$x_1 + x_2 + x_3 + x_4 = a\ , \quad x_5 + x_6 + x_7 + x_8 = 1 - a$$

$$x_1 + x_2 = ba\ , \quad x_1 + x_3 = ca \tag{3.30}$$

and by Atomic Renaming and Obstinacy we add, w.l.o.g.,

$$x_5 = x_6 = x_7 = x_8 = \frac{1-a}{4} \tag{3.31}$$

so, from now on, these are treated as constraints of $K_{a,b,c}$. $Ind(a, b, c)$ is given by

$$Ind(a, b, c) = \left(bca, b(1-c)a, (1-b)ca, (1-b)(1-c)a, \frac{1-a}{4}, \ldots \frac{1-a}{4}\right) \tag{3.32}$$

Let $\vec{x}_{\mathbf{4}}$ denote (x_1, x_2, x_3, x_4). Observe that $Ind(a, b, c)_{\mathbf{4}} = a.Ind(b, c)$. Also, by (3.31), $Mmx(a, b, c)_i = Ind(a, b, c)_i$ for $i = 5, 6, 7, 8$ so the square distance between $Mmx(a, b, c)$ and $Ind(a, b, c)$ is that between $Mmx(a, b, c)_{\mathbf{4}}$ and $Ind(a, b, c)_{\mathbf{4}}$. By a transformation of the constraints we show the following fact:

Claim

$$Mmx(a, b, c)_{\mathbf{4}} = a.Mmx(b, c) \tag{3.33}$$

Proof of claim By Corollary 43, when we calculate $Mmx(a, b, c)$ the constant co-ordinates 5,6,7,8 can be ignored, treating the constraints on x_1, x_2, x_3, x_4 as a constraint set with constant sum a. Let, for $a > 0$, $y_i = x_i/a$ for $i = 1, 2, 3, 4$. Now picking out the *minimax*-best (y_1, y_2, y_3, y_4) is equivalent to picking the *minimax*-best (x_1, x_2, x_3, x_4). The constraints on the y_i's are now exactly $K_{b,c}$. Then taking $\vec{x}_{\mathbf{4}} = a\vec{y}_{\mathbf{4}}$ gives $Mmx(a, b, c)_{\mathbf{4}} = a.Mmx(b, c)$, and we have proved the claim.

■

Using the claim, $|Ind(a, b, c) - Mmx(a, b, c)| = a.|Ind(b, c) - Mmx(b, c)|$ so the largest value of this distance, by Theorem 46, is $\frac{2}{9}$. This value occurs when $a = 1$ and $(b, c) = (\frac{1}{3}, \frac{1}{3}) \ldots$or$(\frac{2}{3}, \frac{2}{3})$. We have proved Theorem 47.

■

Proof of Theorem 44 This follows from Theorem 47.

■

3.1.8 Relativisation

Theorem 48 *Minimax satisfies Relativisation.*

Proof By [[Moh], pp.40-42], the *Renyi Processes* satisfy Relativisation. If

$$K_1 = \{Bel(\phi) = c\} \; + \; \left\{ \sum_{j=1}^{s} a_{ji} Bel(\theta_j|\phi) = b_i \,|\, i = 1, \ldots m \right\}$$
$$K_2 = K_1 \; + \; \left\{ \sum_{j=1}^{q} e_{ji} Bel(\psi_j|\neg\phi) = f_i \,|\, i = 1, \ldots t \right\}, \qquad (3.34)$$

where $0 < c < 1$ and K_1, K_2 are consistent, then for every $\theta \in SL$, $r > 1$, $Ren_r(K_1)(\theta|\phi) = Ren_r(K_2)(\theta|\phi)$. Taking the limit of both sides as $r \to \infty$ gives $Mmx(K_1)(\theta|\phi) = Mmx(K_2)(\theta|\phi)$ and we have proved the theorem.

$\blacksquare$

3.1.9 Irrelevant Information

Although this is a very desirable property of inference processes, it is very rarely satisfied.

Theorem 49 *Minimax does not satisfy Irrelevant Information.*

Proof Let $L = \{p_1, p_2, p_3\}$ and $K_1 = \{Bel(p_1 \wedge p_2) = Bel(p_1 \wedge \neg p_2) + \frac{1}{10}\}$. Then define $K_2 = \{Bel(p_3) = \frac{1}{5}\}$ and let $K = K_1 + K_2$. Clearly if *Minimax* satisfies Irrelevant Information, $Mmx(K)(\theta) = Mmx(K_1)(\theta)$ when θ only mentions p_1 and p_2.

The overlying languages used for each knowledge base are not yet fixed; by Language Invariance (Theorem 39) this is not necessary. We now use the language $\{p_1, p_2\}$ to calculate $Mmx(K_1)$. The four atoms are enumerated $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in the standard ordering with $x_i = Bel(\alpha_i)$ and the constraints are

$$x_1 + x_2 + x_3 + x_4 = 1 \qquad (3.35)$$
$$x_1 - x_2 = \frac{1}{10} \qquad (3.36)$$

By Atomic Renaming and Obstinacy the constraint

$$x_3 = x_4 \tag{3.37}$$

is also true at $Mmx(K_1)$. The solutions of K_1 for which that holds are those

$$\vec{x} = \left(\tau + \frac{2}{20}, \tau, \frac{9}{20} - \tau, \frac{9}{20} - \tau\right) \text{ for which } 0 \leq \tau \leq \frac{9}{20} \tag{3.38}$$

The solution with the least maximum is $(\frac{11}{40}, \frac{7}{40}, \frac{11}{40}, \frac{11}{40})$, since $40\tau > 7$ implies that $\tau + \frac{1}{10} > \frac{11}{40}$, giving a larger maximum and $40\tau < 7$ gives $\frac{9}{20} - \tau > \frac{11}{40}$.

To find $Mmx(K)$ the 8 atoms of L are enumerated in the standard ordering with $x_i = Bel(\alpha_i)$ and the equations are

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 &= 1 & (3.39)\\ x_1 + x_3 + x_5 + x_7 &= \frac{1}{5} & (3.40)\\ x_1 + x_2 - x_3 - x_4 &= \frac{1}{10} & (3.41) \end{aligned}$$

Consider the solution $\vec{p} = \frac{1}{20}(2, 4, 0, 4, 1, 4, 1, 4)$. Note that in general since $x_2 + x_4 + x_6 + x_8 = \frac{16}{20}$ the maximum of x_2, x_4, x_6 and x_8 must be at least $\frac{4}{20}$. Hence $\vec{p}$ has the least possible maximum. If the maximum of a solution of K is $\frac{4}{20}$ then each of those atoms not satisfying p_3 must get belief $\frac{4}{20}$. Assuming that condition, $x_1 = x_3 + \frac{2}{20}$ so $x_1 \geq \frac{2}{20}$ and $x_3 + x_5 + x_7 = \frac{4}{20} - x_1 \leq \frac{2}{20}$ so the maximum of the other co-ordinates could be no lower than $\frac{2}{20}$. In that case, we must have $x_1 = \frac{2}{20}$ and $x_3 = 0$. This leaves x_5 and x_7 which by Renaming can be assumed to be equal, in which case $\vec{p}$ must equal $Mmx(K)$. Hence

$$Mmx(K) = \frac{1}{20}(2, 4, 0, 4, 1, 4, 1, 4) \tag{3.42}$$

Finally, we see that $Mmx(K)(p_1 \wedge p_2) = \frac{3}{10}$ but $Mmx(K_1)(p_1 \wedge p_2) = \frac{11}{40}$ so $Minimax$ does not satisfy Irrelevant Information and we have proved the theorem.

■

3.2 Piecewise Linear Loaf Continuity of $Minimax$

By Theorem 40, $Minimax$ is not continuous. In this section, circumstances are defined in which $Minimax$ does vary continuously. We assume some material from Chapter 6 in this section: in particular Algorithm 94, the $Minimax$ Calculation Algorithm. However, there is no mathematical circularity.

Definition A *loaf* is a set of consistent knowledge bases $\hat{K} = \{K_\lambda \,|\, a \le \lambda \le b\}$ in which each *slice* K_λ is given by

$$\left\{ \sum_{i=1}^{J} a_{ji}x_i = b_j + c_j\lambda \,|\, j = 1, \ldots s \right\} \tag{3.43}$$

for some real numbers $s, a, b, a_{ji}, b_j, c_j$, that are independent of λ. The λ-*range* is $[a, b]$. These definitions apply to the entire thesis.

Theorem 50 *If $\hat{K}$ is a loaf then K_λ is a continuous function of λ in the Blaschke topology.*

Proof Essentially this proof involves making a lot of assumptions about $\hat{K}$ w.l.o.g. and then using Theorem 24.

If all of the K_λ are equal, then Theorem 50 is trivial. Otherwise each knowledge base of the loaf is of the form (3.43). Now we can reparametrise the λ and obtain a loaf $\hat{K'}$ which has λ-range $[0, 1]$ and its slices given by

$$K'_\lambda = \left\{ \sum_{i=1}^{J} a_{ji}x_i = b_j + c_j(a + (b - a)\lambda) \,|\, j = 1, \ldots s \right\} \tag{3.44}$$

We see that K'_λ is equivalent to $K_{a+(b-a)\lambda}$ and if K'_λ is a continuous function of λ in the Blaschke topology, so is K_λ. Hence we assume w.l.o.g. in this proof that $\hat{K}$'s λ-range is $[0, 1]$, so $a = 0, b = 1$.

Also we can rearrange the constraints of $\hat{K}$ so that only one of them depends on λ. If more than one of them has a non-zero coefficient of λ, use one of them to express λ in terms of the x_i's, i.e.

$$\sum_{i=1}^{J} a_{ji}x_i = b_j + c_j\lambda \text{ becomes } \lambda = \sum_{i=1}^{J} \frac{a_{ji}}{c_j}x_i - \frac{b_j}{c_j} \tag{3.45}$$

Then substitute this into all of the other constraints where λ appears, letting the new loaf be $\hat{K}'$. Since $K'_\lambda \equiv K_\lambda$ for every λ such that $\lambda \in [0,1]$, we now assume w.l.o.g. for proving the theorem that $a = 0$, $b = 1$ and only one c_j is non-zero, which equals 1. We can do this by dividing that constraint throughout by the c_j if necessary.

Now each slice is given by

$$K_\lambda = K + \sum_{i=1}^{J} a_i x_i = g + \lambda \tag{3.46}$$

where K is some fixed knowledge base.

We now have assumed enough about the form of $\hat{K}$ and, by using Theorem 24, we see that

$$K_\lambda \to K_0 \quad \text{as} \quad \lambda \searrow 0 \tag{3.47}$$

To get left or right continuity at $\lambda = \lambda_0$ consider a loaf $\hat{K}$ with λ-range $[0, \lambda_0]$ or $[\lambda_0, 1]$ and by a substitution in the constraints get a loaf $\hat{K}'$ with λ-range $[0,1]$ such that either $K'_\lambda \equiv K_{(1-\lambda)\lambda_0}$ for every $\lambda \in [0,1]$, or $K'_\lambda \equiv K_{\lambda_0+\lambda(1-\lambda_0)}$ for every $\lambda \in [0,1]$. Then right continuity at 0 in $\hat{K}'$ implies left or right continuity at λ_0 in $\hat{K}$ as required to complete the proof of the theorem.

■

Definition An inference process N^L is *Piecewise Linear Loaf Continuous* if for every loaf $\hat{K}$ the function $N^L(K_\lambda) : [a,b] \to \mathbb{R}^J$ is continuous and piecewise linear, i.e. there exist $\alpha_0 = a < \alpha_1 < \alpha_2 \ldots < \alpha_t < b = \alpha_{t+1}$ such that $N^L(K_\lambda) : [\alpha_i, \alpha_{i+1}] \to \mathbb{R}^J$ is linear for $i = 0, \ldots t$. This definition applies to the entire thesis.

Remark We can consider Loaf Continuity (that is, without necessarily piecewise linear behaviour) to be a particularly desirable form of continuous behaviour of an inference process, because the slices of a loaf are knowledge bases that only differ in the constants on the right hand sides of the constraints.

To regard Piecewise Linear Loaf Continuity as desirable, we can regard linear behaviour of our inferences as convenient, due to the relative perceived awkwardness of judging the values of nonlinear functions, when we regard inference as a subconscious mental process. This is similar to the remark on page 173—indeed, there may be a strong link between the concepts of Partial Linearity (see Chapter 7) and Piecewise Linear Loaf Continuity. However MD is an example of a non-PL inference process that satisfies PLLC—see Lemma 54, Theorem 142 and the fact that MD is continuous ([Moh]).

Theorem 51 *Minimax is Piecewise Linear Loaf Continuous.*

Proof Consider a loaf $\hat{K}$ as in (3.43). We prove that $Mmx(K_\lambda)$ is a continuous and piecewise linear function of λ. Assume w.l.o.g. that $a = 0, b = 1$, with essentially the same justification as in the proof of Theorem 50. We first prove that $Mmx(K_\lambda)$ is right continuous at $\lambda = 0$ and is linear in a right-neighbourhood of 0.

We now use the $Minimax$ Calculation Algorithm: see Algorithm 94 in Chapter 6. When the algorithm is used with input K_λ, the rank of the constraints is independent of λ and so are the various choices of extra constraints at Step 1. Whether a particular set of extra constraints produces a system of rank J or not is independent of λ. Obtaining $Mmx(K_\lambda)$ involves pre-multiplying the right hand sides of the constraints (which are linear functions of λ) by the inverses of those left hand side matrices (indept. of λ) to get a finite set of linear functions of λ, say

$$\vec{p_1} + \vec{g_1}\lambda, \vec{p_2} + \vec{g_2}\lambda, \ldots \vec{p_q} + \vec{g_q}\lambda \tag{3.48}$$

where $\vec{p_i}, \vec{g_i} \in \mathbb{R}^J$. At Step 2 we remove those containing negative numbers, before comparing the rest compared by least *max* etc, until one remains in Steps 3 and 4; these steps are dependent on λ.

However, when we compare two numbers $A(\lambda) = p_{i_j} + g_{i_j}\lambda$ and $B(\lambda) = p_{d_e} + g_{d_e}\lambda$ for small positive λ

- If $p_{i_j} > (<) p_{d_e}$, $A > (<) B$ for all λ s.t. $0 < \lambda < \delta$ for some δ.
- If $p_{i_j} = p_{d_e}$, then for all positive λ $A > (<) B \iff g_{i_j} > (<) g_{d_e}$.

Since we are comparing a finite number of expressions, there exists $\delta > 0$ such that for all $\lambda \in (0, \delta)$ the truth of all statements of the form $A(\lambda) \leq B(\lambda)$ are constant.

Hence for some i and all sufficiently small λ,

$$Mmx(K_\lambda) = \vec{p_i} + \vec{g_i}\lambda, \tag{3.49}$$

which will be now called $\vec{p} + \vec{g}\lambda$.

Now we return to proving that $Mmx(K_\lambda)$ is right continuous at $\lambda = 0$. It is sufficient that $Mmx(K_0) = \vec{p}$, shown in the lemma below. Suppose that $Mmx(K_0) = \vec{b}$. Now assume that K_0 admits the identity permutation w.r.t. Theorem 28. Then

$$b_1 \geq b_2 \geq \ldots \geq b_J. \tag{3.50}$$

Since $K_\lambda \to K_0$ as $\lambda \to 0$, $\vec{p} \in V^L(K_0)$. Also recall that $m_1 : CL \to \mathbb{R}$ is continuous: see Lemma 31.

Lemma 52 *For each $i = 1, 2, \ldots J$, $p_i = \tilde{p}_i = b_i$.*

Proof This is by induction on i.

Base Case $\underline{i = 1}$ By continuity of m_1 ($= max$), $max(Mmx(K_\lambda)) \to max(Mmx(K_0)) = b_1$ as $\lambda \to 0$. The max of the limit is the limit of the max (max is continuous) so $\tilde{p}_1 = b_1$ and, by (3.50), $p_1 = b_1$.

Inductive Step We assume (I.H.) that the lemma is true for every i not greater than k. Suppose for contradiction that $\tilde{p}_{k+1} \neq b_{k+1}$, then $\tilde{p}_{k+1} > b_{k+1}$ by definition of $Minimax$. Since $\vec{b} - \vec{p}$ is parallel to the hyperplane $\vec{G}_L(K_0)$, it is also parallel to $\vec{G}_L(K_\lambda)$ for all λ since the left hand sides of the constraints are independent of λ, so $\vec{b} + \lambda\vec{g} \in \vec{G}_L(K_\lambda)$.

Now for small enough λ,

$$\vec{y}(\lambda) = \vec{p} + \frac{1}{2}(\vec{b} - \vec{p}) + \lambda\vec{g} \tag{3.51}$$

is non-negative so is a solution of K_λ. To see this we consider a specific co-ordinate, say the i'th. We see that if either $g_i \geq 0$ or $b_i \geq p_i > 0$ then $y_i(\lambda) \geq 0$ for every λ. If $g_i < 0$ and $b_i < p_i$, then if $\lambda < -\frac{b_i}{g_i}$ we see that $y_i(\lambda) \geq 0$. Finally if $b_i = 0$ and $g_i < 0$, $\lambda < \frac{-p_i}{2g_i}$ is sufficient for non-negativity. Taking the minimum of this finite list of upper bounds (across $i = 1, \ldots J$) gives a bound for λ below which $\vec{y}(\lambda) \in V^L(K_\lambda)$.

Hence, by convexity of $V^L(K_\lambda)$, for these values of λ,

$$\vec{w}(\alpha,\lambda) = \vec{p} + \alpha(\vec{b} - \vec{p}) + \lambda\vec{g} \in V^L(K_\lambda) \tag{3.52}$$

for every α s.t. $0 \le \alpha \le \frac{1}{2}$. Let $\vec{z} = \vec{p} + \lambda\vec{g} = \vec{w}(0,\lambda)$. Now by I.H. $p_1 = b_1, \dots p_k = b_k$ so $z_1 = y_1, \dots z_k = y_k$.

Since $p_i = b_i$ for $i = 1, 2, \dots k$, observe that $w_i = z_i$ for each i s.t. $i \le k$. Recall Corollary 43: when comparing $\vec{w}$ and $\vec{z}$ in the *minimax* ordering we can equivalently compare $\vec{W} = (w_{k+1}, \dots w_J)$ and $\vec{Z} = (z_{k+1}, \dots z_J)$.

Let $\lambda = 0$ and $\alpha = 0$. Then since $\tilde{p}_{k+1} > b_{k+1}$, $\tilde{Z}_1 > \tilde{W}_1$. By continuity of *max* this remains true for small values of λ and α. However this contradicts the fact that $\vec{z} = Mmx(K_\lambda)$. Hence the assumption that $\tilde{p}_{k+1} > b_{k+1}$ is false and we have proved the lemma.

$\blacksquare$

Proof of Theorem 51 ctd. We have shown that $Mmx(K_\lambda)$ is right continuous at $\lambda = 0$ and linear in a right neighbourhood of 0. To prove that the function $Mmx(K_\lambda)$ is continuous (on the left or the right) at other points in [0,1] and linear in left/right neighbourhoods of those points, we can use some linear transformations of the loaf similarly to the end of the proof of Theorem 50.

Now we will show the theorem by proving that the local pieces of linear behaviour of $Mmx(K_\lambda)$ join up so that the λ-range ([0,1] w.l.o.g.) is the union of finitely many linear segments as stated.

If there are infinitely many values of λ, λ_i, where the rates of change of $Mmx(K_\lambda)$ above and below $\lambda = \lambda_i$ do not match then by sequential compactness of [0,1] there exists $u \in [0,1]$ s.t. there exist λ_i's $\neq u$ arbitrarily close to it on one side (say above u w.l.o.g.). However by linearity of $Mmx(K_\lambda)$ in a right-neighbourhood of u there exists $\delta > 0$ such that $Mmx(K_\lambda) = \vec{r} + \lambda\vec{s}$ for all $\lambda \in (u, u+\delta)$.

By our assumption, there exists an $\lambda_i \in (u, u+\delta)$. However this is a contradiction since the rate of change of $Mmx(K_\lambda)$ is the same immediately above and below that λ_i, i.e. $\vec{s}$. Hence there are only finitely many λ values where the direction of *Minimax* changes, as required, and we have proved the theorem.

■

3.3 Is Piecewise Linear Loaf Continuity a property of the *Renyi Processes*, or of CM_∞?

Theorem 53 *Ren_2 (i.e. MD) satisfies Piecewise Linear Loaf Continuity. No other Renyi Process Ren_r satisfies this property. Neither does CM_∞.*

Lemma 54 *Ren_2 satisfies Piecewise Linear Loaf Continuity.*

Proof Similarly to the proof of Theorem 51, it is sufficient to prove, for every loaf $\hat{K}$ for which the λ-range is $[0,1]$, that $Ren_2(K_\lambda)$ is right continuous at 0 and is linear in a right-neighbourhood of 0. By [Moh], we know that Ren_2 is continuous.

Notation A *Z-constraint* is a constraint of the form $(x_i = 0)$ for some i. This definition is used throughout the thesis.

We now introduce, just for the proof of Lemma 54, technical definitions and results about convex polytopes.

Definition Let some subset F of a convex polytope P be expressible in the form

$$F = \{\vec{x} \in P \text{ s.t. } a_1x_1 + x_2x_2 + \ldots + a_Jx_J = a_0\} \tag{3.53}$$

where $a_0, a_1, \ldots a_J$ are constants and $a_1x_1 + a_2x_2 + \ldots + a_Jx_J \geq a_0$ for all $\vec{x} \in P$. Then F is a *face* of P. Note that this definition is equivalent to that of "face" in [Zie], in which the author uses $\leq$ instead of our $\geq$. Also if $F \neq P$, F is a *proper face* of P.

Claim Every point, $\vec{x}$, on a proper face of P defined by $\sum_{i=1}^J a_ix_i = a_0$ as above, is on the topological boundary of P, using the subspace topology of P.

Proof of claim This is because there exists a point, say $\vec{y}$, in P for which $\sum_{i=1}^J a_iy_i > a_0$ so, as $\epsilon \searrow 0$, $\vec{x} + \epsilon(\vec{x} - \vec{y}) \to \vec{x}$ and is not in P, but these points do lie in the smallest affine set which contains P.

■

In [Zie], on page 51, the author defines a "vertex" of a convex polytope P in a different way to the definition in this thesis: he defines a vertex of P to be a face of P which consists of a single point.

The following lemma shows that we can refer to results from [Zie] in the following proofs, without ambiguity, as the author's definition of a vertex is equivalent to the definition we use in this thesis.

Lemma 55 ([Zie], page 52) *Let P be a convex polytope. Then the vertices of P (as given by* (1.15)*) are precisely those points $\vec{x}$ of P for which $\{\vec{x}\}$ is a face of P.*

Notation For all $K \in CL$ and each $\{i_1, \dots i_d\} \subseteq \{1, \dots J\}$, let $F(K)(i_1, i_2, \dots i_d)$ denote $V(K + \{x_{i_s} = 0 \text{ for } s = 1, 2, \dots d\})$ [1]. Also let $G(K)(i_1, i_2 \dots i_d)$ denote $G(K + \{x_{i_s} = 0 \text{ for } s = 1, 2, \dots d\})$.

Sublemma 56 *For any $K \in CL$, the faces of $V(K)$ are precisely the sets of the form $F(K)(i_1, \dots i_d)$.*

Proof Firstly we show that each set of the form $F(K)(i_1, \dots i_d)$ is a face of $V(K)$. For each $i_1 = 1, \dots J$, $V(K + x_{i_1} = 0)$ is a face of $V(K)$ since $x_{i_1} \geq 0$ for every $\vec{x} \in V(K)$. By [[Zie], page 53], each face of a face of $V(K)$ is a face of $V(K)$ so if we add a set of Z-constraints to K, the solution set of the resulting knowledge base (which is not necessarily consistent) is a face of K.

Let F be a face of $V(K)$. If $F = V(K)$, let $\{i_1, \dots i_d\} = \emptyset$ and we are done (see [1]). Otherwise, let F be a proper face of $V(K)$. By the above claim, F is on the topological boundary of $V(K)$ so, for every $\vec{x} \in F$, there exists a co-ordinate i such that $x_i = 0$ but $y_i > 0$ for some $\vec{y} \in V(K)$. By convexity of F, there exists a co-ordinate i_1 such that $x_{i_1} = 0$ for all $\vec{x} \in F$ and such that $y_{i_1} > 0$ for some $\vec{y} \in V(K)$. Hence

$$F \subseteq F(K)(i_1) = V(K + x_{i_1} = 0) \subseteq V(K) \tag{3.54}$$

Either the above is an equality, so we are done, or F is a proper face of $F(K)(i_1)$, by [[Zie], page 53], so there exists an i_2 such that

$$F \subseteq F(K)(i_1, i_2) = V(K + x_{i_1} = 0, x_{i_2} = 0) \subseteq F(K)(i_1) \subseteq V(K) \tag{3.55}$$

[1] If $\{i_1, \dots i_d\} = \emptyset$, $F(K)(..) = V(K)$ and $G(K)(..) = G(K)$, where $(..)$ denotes $(i_1, \dots i_d)$ in the case $d = 0$.

and so on. This process must stop since $V(K + x_1 = 0, x_2 = 0, \ldots x_J = 0) = \emptyset$. Hence we have proved the sublemma.

■

Sublemma 57 $V(K)$ *is the disjoint union of the interiors of its faces, where the subspace topology relative to each face is used to define its interior.*

Proof See [[Zie], page 61].

■

Proof of Lemma 54 continued For all $\lambda \in [0, 1]$, $\vec{X} = MD(K_\lambda)$ lies in the interior of exactly one face of $V(K_\lambda)$, by Sublemma 57. Let this face be given by $F{*} = F(K)(i_1, \ldots i_d)$. We assume w.l.o.g. that $G{*} = G(K)(i_1, \ldots i_d)$ really is the smallest affine hyperplane containing $F{*}$. If that is not true, choose $\vec{y}$ in the interior of $F{*}$ and extend $i_1, \ldots i_d$ to include every co-ordinate i such that $y_i = 0$.

Now, for each direction $\vec{dir}$ parallel to $G{*}$, $dir_{i_1} = 0, \ldots dir_{i_d} = 0$ so moving a short enough distance in this direction, from $\vec{y}$, will keep every co-ordinate value non-negative. Hence every direction parallel to $G{*}$ is also parallel to $V{*}$. This implies that $\vec{X}$ is in the interior of $V{*}$ relative to the topology of $G{*}$. The derivative of $\sum_{i=1}^{J} x_i^2$ at $\vec{X}$ in every direction parallel to $G{*}$ is zero so by doing the differentiation we see that

$$\vec{X} \cdot \vec{dir} = 0 \tag{3.56}$$

for all $\vec{dir}$ parallel to $G{*}$, which is a linear equation in $\vec{X}$.

Suppose that $\{i_1, \ldots i_d\}$ is some subset of $\{1, \ldots J\}$. Then if, for some λ, $MD(K_\lambda)$ is in the interior of $F{*}$ as above and we consider the constraints of $K_\lambda + \{x_{i_1} = 0, x_{i_2} = 0, \ldots x_{i_d} = 0\}$, the left hand sides are independent of λ and right hand sides are linear functions of λ; this argument is similar to the beginning of the proof of Theorem 51.

Hence the generalised solutions take the form

$$G^L(K_\lambda + \{x_{i_1} = 0, \ldots x_{i_d} = 0\}) = \left\{ \vec{p} + \vec{g}\lambda + \sum_{i=1}^{d} z_i \vec{u_i} \text{ s.t. } z_1, z_2, \ldots z_d \in \mathbb{R} \right\} \tag{3.57}$$

for some $\vec{p}, \vec{g} \in \mathbb{D}^J$, $\vec{u_i}$'s $\in \mathbb{R}^J$ that are independent of λ. We assume w.l.o.g. that the closest point to $(\frac{1}{J}, \ldots \frac{1}{J})$ in the hyperplane is $\vec{p} + \vec{g}\lambda$ for all λ; we can do this because the equation (3.56) is linear in $\vec{X}$.

For each $S = \{i_1, \ldots i_d\} \subseteq \{1, \ldots J\}$, we obtain $\vec{p}(S) + \vec{g}(S)\lambda$ which is a possible value of $MD(K_\lambda)$, using the method above. For each particular value of λ, $MD(K_\lambda)$ must equal one of these possibilities; i.e. the one which is closest to $(\frac{1}{J}, \frac{1}{J}, \ldots \frac{1}{J})$. If two possibilities, $\vec{p}(S) + \vec{g}(S)\lambda$ and $\vec{p}(S') + \vec{g}(S)\lambda$ are distinct, either

- For all $\lambda \in \mathbb{R}$ they are equally close to $(\frac{1}{J}, \ldots \frac{1}{J})$. In this case, neither value can equal MD or
- They are equally close to $(\frac{1}{J}, \ldots \frac{1}{J})$ for at most two values of λ.

Therefore there exists $\epsilon > 0$ such that there exists $S* \subseteq \{1, \ldots J\}$ such that $MD(K_\lambda) = \vec{p}(S*) + \vec{g}(S*)\lambda$ for all $\lambda \in (0, \epsilon)$. Hence we have shown linearity of $MD(K_\lambda)$ in a right-neighbourhood of $\lambda = 0$. We can now continue in a similar way to the final part of the proof of Theorem 51 to complete the proof of Lemma 54.

$\blacksquare$

Lemma 58 *The Renyi Processes Ren_r for which $0 < r < 1$ do not satisfy Piecewise Linear Loaf Continuity.*

Proof We consider a particular loaf and show that Piecewise Linear Loaf Continuity must fail. For the language $L = \{p_1, p_2\}$ we use the standard ordering of the four atoms of L. For all λ such that $0 \leq \lambda \leq \frac{1}{3}$, let

$$K_\lambda = \left\{ x_1 = 0, \sum_{i=1}^{4} x_i = 1, x_3 + 2x_2 = \lambda \right\} \tag{3.58}$$

where $x_i = Bel(\alpha_i)$ for $i = 1, 2, 3, 4$ as usual. $\hat{K}$ is the loaf consisting of the slices K_λ as above with λ-range $[0, 1/3]$. Now

$$V^L(K_\lambda) = \left\{ (0, \tau, \lambda - 2\tau, 1 - \lambda + \tau) | 0 \leq \tau \leq \frac{\lambda}{2} \right\} \tag{3.59}$$

and define $\vec{x}(\lambda, \tau) = (0, \tau, \lambda - 2\tau, 1 - \lambda + \tau)$. We let $r = R$ be a fixed real number in $(0, 1)$. For all $\lambda \in [0, \frac{1}{3}]$ define $X(\lambda)$ to be the value of τ such that

$$Ren_R(K_\lambda) = \vec{x}(\lambda, \tau) = (0, \tau, \lambda - 2\tau, 1 - \lambda + \tau) \tag{3.60}$$

Now $X(0) = 0$ since $\vec{x}(0, 0)$ is the only solution of K_0. For all K_λ, $Ren_R(K_\lambda)$ is the solution for which $\sum_{i=1}^{4} x_i^R$ is maximal. It is now useful for us to look at how the sign of the derivative w.r.t. τ of this quantity varies as τ increases from 0 to $\lambda/2$. That derivative is

$$R(\tau^{R-1} - 2(\lambda - 2\tau)^{R-1} + (1 - \lambda + \tau)^{R-1}) \tag{3.61}$$

This tends to ∞ as $\tau \to 0$ and tends to $-\infty$ as $\tau \to \lambda/2$. Using the fact that $\sum_{i=1}^{4} x_i^R$ is a concave function (see [Moh]) with a negative second derivative from all points in $\mathbb{D}^4$ in all directions parallel to $\mathbb{D}^4$, we know that for all $\lambda \in (0, \frac{1}{3})$, $X(\lambda)$ is the unique value of τ such that

$$\tau^{R-1} - 2(\lambda - 2\tau)^{R-1} + (1 - \lambda + \tau)^{R-1} = 0 \tag{3.62}$$

lying in $(0, \lambda/2)$. We denote $\frac{dX}{db}$ by X'. If the above equation, with $X(\lambda)$ in place of τ, is differentiated implicitly w.r.t. λ we obtain

$$X'X^{R-2} - 2(1 - 2X')(\lambda - 2X)^{R-2} + (X' - 1)(1 - \lambda + X)^{R-2} = 0 \tag{3.63}$$

for all $\lambda \in (0, 1/3)$. The coefficient of X' in the above equation is

$$X^{R-2} + 4(\lambda - 2X)^{R-2} + (1 - \lambda + X)^{R-2} \tag{3.64}$$

which is always strictly positive, so, by (3.63), X' is a function of X and λ. Since Ren_R is continuous (by [Moh]), $X(\lambda)$ is a continuous function. Using (3.63), we see that $X(\lambda)$ is differentiable with a continuous derivative.

Assume for contradiction that Ren_R satisfies Piecewise Linear Loaf Continuity. Then $X(\lambda)$ is a piecewise linear continuous function and, since it is differentiable, it must be linear. By continuity of Ren_R, $X(\lambda) \to 0$ as $\lambda \searrow 0$. Therefore for some constant c and all $\lambda \in (0, 1/3)$, $X(\lambda) = c\lambda$.

We now substitute $X(\lambda) = c\lambda$ and $X' = c$ into (3.63) to obtain

$$c(c\lambda)^{R-2} - 2(1-2c)(\lambda - 2c\lambda)^{R-2} + (c-1)(1-\lambda+c\lambda)^{R-2} = 0 \tag{3.65}$$

for all $\lambda \in (0, 1/3)$. This can be rearranged, giving

$$\left(\frac{1-c}{c^{R-1} - 2(1-2c)^{R-1}}\right) = \left(\frac{\lambda}{1-\lambda+c\lambda}\right)^{R-2} \tag{3.66}$$

Even if the left hand side is not well defined, it is independent of λ. The right hand side definitely takes real values that are different for different values of λ so we have a contradiction and the lemma is proved.

■

Lemma 59 *The Renyi Processes Ren_r for which $r > 1$ and $r \neq 2$ do not satisfy Piecewise Linear Loaf Continuity.*

Proof Similarly to the proof of Lemma 58, we let $r = R$ be a real number greater than 1 and not equal to 2, and define the slices of a loaf, $\hat{K}$, to be

$$K_\lambda = \left\{x_1 = 0, \sum_{i=1}^{4} x_i = 1, x_3 + 2x_2 = \lambda\right\} \tag{3.67}$$

but this time the λ-range is $[\lambda_0, 1]$, where

$$\lambda_0 = \frac{1}{1 + 2^{\frac{1}{R-1}}} \tag{3.68}$$

As for Lemma 58, we define $\vec{x}(\lambda, \tau) = (0, \tau, \lambda - 2\tau, 1 - \lambda + \tau)$ and
$V^L(K_\lambda) = \{\vec{x}(\lambda, \tau) \ \text{ s.t. } \ 0 \le \tau \le \lambda/2\}$. For all $\lambda \in [\lambda_0, 1]$, we let $X(\lambda)$ be the value of τ such that $Ren_R(K_\lambda) = \vec{x}(\lambda, \tau)$ and for contradiction we assume that $X(\lambda)$ is a piecewise linear function of λ.

For all λ, $Ren_R(K_\lambda)$ is the solution of K_λ for which $\sum_{i=1}^{4} x_i^R$ is minimal. This quantity, and its derivative w.r.t. τ, have the same formulae as before—see (3.61). We now inspect the sign of the derivative at $\tau = 0$ and $\tau = \lambda/2$:

At $\tau = 0$, the derivative is $R(-2\lambda^{R-1} + (1-\lambda)^{R-1})$ which is zero iff $\left(\frac{1-\lambda}{\lambda}\right)^{R-1} = 2$ iff $\lambda = \lambda_0$. This explains the choice of λ-range (3.68), which may otherwise seem strange! If $\lambda > \lambda_0$, the derivative at $\tau = 0$ is negative.

At $\tau = \lambda/2$, the derivative is $R((\frac{\lambda}{2})^{R-1} + (1-\frac{\lambda}{2})^{R-1}) > 0$ for all $\lambda \in [\lambda_0, 1]$.

The second derivative of $\sum_{i=1}^{4} x_i^R$ is positive from all points in $\mathbb{D}^4$ along all directions parallel to $\mathbb{D}^4$ (it is convex, see [Moh]) so we know that $X(\lambda_0) = 0$ and, for all $\lambda \in (\lambda_0, 1]$, $X(\lambda)$ is the unique value of τ for which

$$\tau^{R-1} - 2(\lambda - 2\tau)^{R-1} + (1-\lambda+\tau)^{R-1} = 0 \tag{3.69}$$

and $\tau \in (0, \frac{\lambda}{2})$. If we let

$$K = \{x_1 = 0, x_2 + x_3 + x_4 = 0\} \tag{3.70}$$

then $Ren_R(K) = (0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ by Atomic Renaming. Since
$K_1 = K + \{x_3 + 2x_2 = 1\}$ is also satisfied by $\vec{x}(1, \frac{1}{3}) = (0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$,

$$Ren_R(K_1) = \left(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \tag{3.71}$$

by Obstinacy of Ren_R, so $X(1) = \frac{1}{3}$.

We now differentiate (3.69) implicity, as for Lemma 58, to obtain (3.63), true for all $\lambda \in (\lambda_0, 1)$. By the same reasoning as before, $X(\lambda)$ must be linear and, by continuity of Ren_R, $X \to 0$ as $\lambda \searrow \lambda_0$ and $X \to \frac{1}{3}$ as $\lambda \nearrow 1$. However, since $X(\lambda_0)$ and $X(1)$ are known, the constant value of $X' = \frac{dX}{d\lambda}$ must be

$$c = \frac{\frac{1}{3} - 0}{1 - \lambda_0} = \frac{1}{3}\left(1 + 2^{-\frac{1}{R-1}}\right) \tag{3.72}$$

We substitute $X' = c$ and $X = \frac{1}{3} + c(\lambda - 1)$ into (3.63), giving

$$\begin{aligned}
&\left(1 + 2^{-\frac{1}{R-1}}\right)\left[\lambda\left(\frac{1}{3} + \frac{2^{-\frac{1}{R-1}}}{3}\right) - \frac{2^{-\frac{1}{R-1}}}{3}\right]^{R-2} \\
&\left(-2 + 2^{\frac{2R-3}{R-1}}\right)\left[\lambda\left(\frac{1}{3} - \frac{2^{\frac{R-2}{R-1}}}{3}\right) + \frac{2^{-\frac{R-2}{R-1}}}{3}\right]^{R-2} \\
+&\left(-2 + 2^{-\frac{1}{R-1}}\right)\left[\lambda\left(-\frac{2}{3} + \frac{2^{-\frac{1}{R-1}}}{3}\right) - \frac{2^{-\frac{1}{R-1}}}{3} + 1\right]^{R-2} = 0
\end{aligned} \tag{3.73}$$

for all $\lambda \in (\lambda_0, 1)$. By continuity we can substitute $\lambda = 1$ into the above equation, giving

$$\left(1 + 2^{-\frac{1}{R-1}}\right)\left[\frac{1}{3}\right]^{R-2} + \left(-2 + 2^{\frac{2R-3}{R-1}}\right)\left[\frac{1}{3}\right]^{R-2} + \left(-2 + 2^{-\frac{1}{R-1}}\right)\left[\frac{1}{3}\right]^{R-2} = 0 \tag{3.74}$$

so that

$$2.2^{-\frac{1}{R-1}} + 2^{2R-3}R - 1 - 3 = 0 \tag{3.75}$$

$$\Rightarrow \quad 6.2^{-\frac{1}{R-1}} = 3 \quad \Rightarrow \quad 2^{-\frac{1}{R-1}} = 2^{-1} \quad \Rightarrow \quad R = 2 \tag{3.76}$$

Since $R \neq 2$, by our assumption, we have deduced a contradiction and have proved the lemma.

■

Lemma 60 *Maximum Entropy does not satisfy Piecewise Linear Loaf Continuity.*

Proof We define a loaf $\hat{K}$ to have slices given by

$$K_\lambda = \{Bel(p_1) = \lambda, Bel(p_2) = \lambda\} \tag{3.77}$$

with λ-range $[0, 1/2]$. Then using the standard ordering of the atoms $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ of $L = \{p_1, p_2\}$ and $x_i = Bel(\alpha_i)$ as usual we write

$$K_\lambda = \left\{ \sum_{i=1}^{4} x_i = 1,\ x_1 + x_2 = \lambda,\ x_1 + x_3 = \lambda \right\} \tag{3.78}$$

Let $\vec{x}(\lambda, \tau)$ denote $(\tau, \lambda - \tau, \lambda - \tau, 1 - 2\lambda + \tau)$. Then for all $\lambda \in \left[0, \frac{1}{2}\right]$,

$$V^L(K_\lambda) = \{\vec{x}(\lambda, \tau) \text{ s.t. } 0 \leq \tau \leq \lambda\} \tag{3.79}$$

Now, since *Maximum Entropy* satisfies Independence (by [ParVen1]), $ME(K_\lambda) = \vec{x}(\lambda, \lambda^2)$ for all λ. Since λ^2 is not a piecewise linear function of λ, *Maximum Entropy* does not satisfy Piecewise Linear Loaf Continuity and we have proved the lemma.

■

Lemma 61 *CM_∞ does not satisfy Piecewise Linear Loaf Continuity.*

Proof We use the same loaf $\hat{K}$ as in the previous lemma, so that for all λ such that $0 \leq \lambda \leq 1/2$,

$$V^L(K_\lambda) = \{\vec{x}(\lambda, \tau) \text{ s.t. } 0 \leq \tau \leq \lambda\} \tag{3.80}$$

where $\vec{x}(\lambda,\tau)$ denotes $(\tau, \lambda-\tau, \lambda-\tau, 1-2\lambda+\tau)$. To calculate CM_∞, we need to maximise $\log(\tau)+2\log(\lambda-\tau)+\log(1-2\lambda+\tau)$, so we maximise

$$f(\tau) = \tau(\lambda-\tau)^2(1-2\lambda+\tau) \tag{3.81}$$

Similarly to the previous lemmas, for all $\lambda \in (0, 1/2]$ we let $X(\lambda)$ be the unique value $\tau = X$ such that $f(\tau)$ is maximal.

Assume for contradiction that CM_∞ satisfies Piecewise Linear Loaf Continuity. Then $X(\lambda)$ is a piecewise linear function of λ. Since $X(0) = 0$, there exists a constant c such that $X(\lambda) = c\lambda$ for all sufficiently small positive values of λ, say all $\lambda < \delta$. CM_∞ satisfies Open-mindedness (by Theorem 21) so

$$X(\lambda) \neq 0, X(\lambda) \neq \lambda \tag{3.82}$$

for all $\lambda > 0$.

Hence, where f' denotes $df/d\tau$, $f'(X(\lambda)) = 0$ for all $\lambda \in (0,\delta)$. When we do the algebra, this implies that

$$\lambda^2 - 2\lambda^3 + 10c\lambda^3 - 4c\lambda^2 + 3c^2\lambda^2 - 12c^2\lambda^3 + 4c^3\lambda^3 = 0 \tag{3.83}$$

if $\lambda \in (0,\delta)$ so

$$3c^2 - 4c + 1 + \lambda(4c^3 - 12c^2 + 10c - 2) = 0 \tag{3.84}$$

which must mean that $3c^2 - 4c + 1 = 4c^3 - 12c^2 + 10c - 2 = 0$. We now know that $c = 1$. By (3.82), we have deduced a contradiction so CM_∞ does not satisfy Piecewise Linear Loaf Continuity and we have proved the lemma.

■

Proof of Theorem 53 Together Lemmas 54, 58, 59, 60 and 61 give the theorem.

■

Chapter 4

The dual of $Minimax$ as a limit of inference processes

The inference process $Minimax$ arises (Chapter 2) from considering the repetition of an experiment whose possible result is one of J mutually exclusive outcomes, the atoms of a language L, which have belief values $\vec{x} = (x_1, \ldots x_J)$. Recall that we suppose that when the experiment is carried out r times we wish to minimise the probability that the same outcome occurs every time. Hence we justify using the *Renyi Process* Ren_r, and by taking the limit as $r \to \infty$, we arrive at $Minimax$.

4.1 Justifying the $Every_n^L$ inference processes

Consider instead the following scenario. The experiment is repeated n times but now the philosophy is that each outcome should have as much chance as possible of occurrence, rather than that no outcome should dominate.

In the work "Theodicee" by Leibniz (1710), the author argued that the universe should contain every possible phenomenon. This fitted into Leibniz' overall view of the universe being, within certain bounds, the best possible universe God could have created. Other philosophers, including Arthur Lovejoy ([Lov]), have also supported the "Plenitude Principle": that everything that can exist, does exist.

We assume that the starting conditions of the experiment are repeatable, from this point of view we should maximise the probability that after r trials every outcome has occurred at least once. We regard each outcome α_i for which the knowledge base forces $Bel(\alpha_i) = 0$ as impossible, in a probabilistic sense. There are some issues to clear up before this can be made a well-defined inference process:

- If some $\alpha \in At^L$ are s.t. $Bel(\alpha) = 0$, the probability of every outcome occurring in the r trials is zero.
- There may not be enough trials to allow every outcome to occur.
- Is the optimal $\vec{x} \in V^L(K)$ unique?

Notation We ignore the outcomes which we know to be impossible and suppose that our knowledge base is $K \in CL$. Hence we use the notation from Section 2.3 and write K in *dezeroed form.* We assume w.l.o.g. that the atoms of L are enumerated $\alpha_1, \ldots \alpha_J$, such that

$$I^L(K) = \{i \text{ s.t. } J' < i \leq J\} \tag{4.1}$$

for some integer J' such that $2 \leq J' \leq J$. We can do this since our work is symmetrical w.r.t. the labelling of the atoms. As usual we let $x_i = Bel(\alpha_i)$ for each $i = 1, \ldots J$. Since all of our work respects equivalence of knowledge bases, we assume (as in Section 2.3) w.l.o.g. that K includes constraints of the form $x_i = 0$ for each i s.t. $J' < i \leq J$ and that every other constraint only refers to $x_1, x_2, \ldots x_{J'}$. Solutions of K are written as points in $\mathbb{D}^{J'}$, ignoring the constant zeros.

We let $All(\vec{x}, r)$ denote the probability that in r identical trials of the experiment, for which the *possible* outcomes of a trial have probabilities $x_1, \ldots x_{J'}$, every outcome happens at least once. We use the multinomial expansion of $(x_1 + \ldots + x_{J'})^r$ to deduce that

$$All(\vec{x}, r) = \sum_{\substack{\vec{p} \in \mathbb{N}^{J'}, \\ \sum_{i=1}^{J'} p_i = r, \text{ and all } p_i > 0}}^{J} \prod x_i^{p_i} \frac{r!}{p_1! \ldots p_{J'}!} \tag{4.2}$$

Definition We define the inference process $Every_r^L$ by

$$Every_r^L(K) = \text{the unique } \vec{x} \in V^L(K) \text{ such that } All(\vec{x}, r) \text{ is maximal} \tag{4.3}$$

when such an $\vec{x}$ is unique, otherwise $Every_r^L(K)$ is undefined.

We now tackle the question: "Given that K is fixed, which r make $Every_r^L(K)$ well-defined?"

When $r \geq J'$ there are enough repetitions of the experiment to make the probability $All(\vec{x}, r) > 0$ for some $\vec{x} \in V^L(K)$ since, by the convexity of $V^L(K)$, there is a solution of K giving strictly positive belief to each $\alpha_1, \dots \alpha_{J'}$—see [[Par], page 74].

Theorem 62 *If $r = J'$ or $r = J' + 1$, ${Every_r}^L(K) = CM_\infty(K)$.*

Proof For $r = J'$ the only $\vec{p}$ in the formula is $(1, 1, 1, \dots 1)$ so, since $\{1, \dots J\} \setminus I^L(K) = \{1, \dots J'\}$,

$$All(\vec{x}, J') = \prod_{i=1}^{J'} x_i \frac{J'!}{1! \dots 1!} = J'!\, 2^{(\sum_{i=1}^{J'} \log(x_i))} \tag{4.4}$$

Hence maximising this is equivalent to maximising $\sum_{i=1}^{J'} \log(x_i)$, which gives $CM_\infty(K)$ by (1.78).

When $r = J' + 1$ the J' possible values of $\vec{p}$'s in the formula for $Every_r$ are a 2 and $J' - 1$ 1's in every possible order.

$$\begin{aligned} & \frac{2\, All(\vec{x}, J'+1)}{J'+1\,!} \\ = \;& (x_1^2 x_2 x_3 \dots x_{J'} + x_1 x_2^2 x_3 x_4 \dots x_{J'} + x_1 x_2 x_3^2 \dots x_{J'} + \dots + x_1 x_2 x_3 \dots x_{J'}^2) \\ = \;& (x_1 + x_2 + \dots + x_{J'})(x_1 x_2 \dots x_{J'}) = \text{ constant } \times \prod_{i=1}^{J'} x_i \qquad (4.5) \end{aligned}$$

since $\sum_{i=1}^{J'} x_i = 1$. Hence $Every_{J'+1}$ is well-defined and also equals CM_∞.

■

Remark Recall Theorem 12, which implies that if $All(\vec{x}, r)$ is *concave*, there exists a unique $\vec{x} \in V^L(K)$ for which $All(\vec{x}, r)$ is maximal. However $All(\vec{x}, r)$ is not necessarily concave when $r \geq J'$. For example consider $J' = 3$, $r = 6$ and let $\vec{x} = (0.8, 0.1, 0.1)$, $\vec{a} = (0.802, 0.099, 0.099)$ and $\vec{b} = (0.798, 0.101, 0.101)$. We see that $\vec{x} = \frac{1}{2}(\vec{a} + \vec{b})$. Calculating from (4.2) gives

$$All(\vec{a}, 6) = 0.196054855, All(\vec{b}, 6) = 0.202354747 \tag{4.6}$$

so $\frac{1}{2}(All(\vec{a},6)+All(\vec{b},6)) = 0.1992048$ to 7 significant figures but $All(\vec{x},6) = 0.1992000$ to 7 s.f., contradicting concavity.

Notation From here on (in this chapter) it is more convenient to use the quantity $Nall$, given by

$$\begin{aligned} Nall(\vec{x},r) &= 1 - All(\vec{x},r) \\ &= \text{the probability that in } r \text{ trials some outcome fails to occur.} \end{aligned} \tag{4.7}$$

In the light of Theorem 12, we desire that $Nall$ be convex so that there is a unique point minimising it within $V(K)$. By Theorem 5, this is true if $Nall$ has a positive second derivative everywhere in $\mathbb{D}^{J'}$; in the statement of Theorem 5, we are using $S = \mathbb{D}^{J'} = \vec{x} \in \mathbb{D}^J$ s.t. $x_i = 0$ for all $i > J'$. We use $V = V^L(K) \subseteq \mathbb{D}^{J'}$. Our dezeroed notation means that we are really considering taking the second derivative of $Nall$ in each direction through $\mathbb{D}^J$ which preserves $x_i = 0$ for all $i \in I^L(K)$.

Although the above remark shows that these derivatives are not always positive, Lemma 63 below shows that $Nall(\vec{x},r)$ is convex, provided that each x_i is bounded away from zero and r is large enough.

By using the Inclusion-Exclusion principle, we deduce that

$$Nall(\vec{x},r) = \sum_{\emptyset \neq S \subset \{1,\ldots J'\}} (-1)^{|S|+1} \left(1 - \sum_{i \in S} x_i\right)^r \tag{4.8}$$

where each term is the probability that every outcome in S fails to occur and we assume that $S \neq \{1,\ldots J'\}$ since the term given by $S = \{1,\ldots J'\}$ contributes zero.

Lemma 63 *For all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $r > N$, $Nall(\vec{x},r)$ is convex when restricted to $\{\vec{x} \in \mathbb{D}^{J'} \,|\, x_i > \epsilon$ for $i = 1,\ldots J'\}$.*

Proof We can assume w.l.o.g. that $\epsilon < 1/J'$. We must show that the second derivative at such $\vec{x}$ is positive in every direction that preserves $\sum_{i=1}^{J'} x_i = 1$ (and that preserves $x_i = 0$ for all $i \in I^L(K)$), i.e. in directions $\vec{s}$ s.t. $\sum_{i=1}^{J'} s_i = 0$. The second derivative in such a direction $\vec{s}$ is given by (see e.g. [[Egg], p 51])

$$Nall''(\vec{x}, \vec{s}) = \frac{1}{r(r-1)} \sum_{i=1}^{J'} \sum_{j=1}^{J'} \frac{\partial^2(Nall)}{\partial x_i \partial x_j} s_i s_j \tag{4.9}$$

Without loss of generality, a typical term of (4.8) is $(-1)^{k+1}(1 - x_1 - x_2 \ldots - x_k)^r$ and, upon being differentiated w.r.t. first x_i and then w.r.t. x_j, the result is

$$r(r-1)(-1)^{k+1}(1 - x_1 - x_2 \ldots - x_k)^{r-2} \tag{4.10}$$

if $\{i, j\} \subseteq \{1, 2, \ldots k\}$, 0 otherwise. Now the coefficient of $(-1)^{k+1}(1 - x_1 - \ldots - x_k)^{r-2}$ in $Nall''$ is

$$\sum_{i=1}^{k} \sum_{j=1}^{k} s_i s_j = \left(\sum_{i=1}^{k} s_i \right)^2 \tag{4.11}$$

and similarly for other subsets S of $\{1, 2, \ldots J'\}$ s.t. $|S| = k$ so

$$Nall''(\vec{x}, \vec{s}) = \sum_{\emptyset \neq S \subset \{1, \ldots J'\}} (-1)^{|S|+1} \left(1 - \sum_{i \in S} x_i \right)^{r-2} \left(\sum_{i \in S} s_i \right)^2 \tag{4.12}$$

The following claim is sufficient to complete the proof of Lemma 63.

Claim Let $N \geq 2 + \frac{(J'+2)}{\log(\frac{1}{1-\epsilon})}$. Let $\vec{x} \in \mathbb{D}^{J'}$ be s.t. $x_i > \epsilon$ for every $i \leq J'$ and let $\vec{s} \in \mathbb{R}^{J'}$ be s.t. $\sum_{i=1}^{J'} s_i = 0$. Then for every $r > N$, $Nall''(\vec{x}, \vec{s}) > 0$.

Proof of claim Let N be as above, and $r > N$. Let $r' = r - 2$ so that $r' > \frac{\log(4.2^{J'})}{\log(\frac{1}{1-\epsilon})}$. Since $\log\left(\frac{x}{x-\epsilon}\right)$ is a decreasing function of x for $x \in (\epsilon, 1]$, if $\epsilon < a \leq 1$, then

$$\begin{aligned} r' &> \frac{\log(4.2^{J'})}{\log(a) - \log(a-\epsilon)} \\ \implies r' \log(a) &> r' \log(a-\epsilon) + \log(4.2^{J'}) \end{aligned} \tag{4.13}$$

$$\implies a^{r'} > 4.2^{J'}(a-\epsilon)^{r'} \tag{4.14}$$

Now there are less than $2^{J'}$ terms in (4.12). It is enough to show that for the largest negative term there is a positive one with modulus more than $2^{J'}$ times bigger, so that it outweighs all of the negative terms. Consider w.l.o.g. the term

$$-(1 - x_1 - x_2 \ldots - x_k)^{r'}(s_1 + s_2 \ldots + s_k)^2 \tag{4.15}$$

where k is even. We now show that one of the positive terms

$(1-\sum_{i=2}^{k} x_i)^{r'}(\sum_{i=2}^{k} s_i)^2, (1-\sum_{i=1,\neq 2}^{k} x_i)^{r'}(\sum_{i=1,\neq 2}^{k} s_i)^2, \ldots$
$(1-\sum_{i=1,\neq k}^{k} x_i)^{r'}(\sum_{i=1,\neq k}^{k} s_k)^2$ is at least $2^{J'}$ times bigger.

If $X = 1 - x_1 - \ldots - x_k$ and $P = s_1 + \ldots + s_k$ the positive terms are given by $(X+x_1)^{r'}(P-s_1)^2, (X+x_2)^{r'}(P-s_2)^2, \ldots (X+x_k)^{r'}(P-s_k)^2$ and the negative term is $-X^{r'}P^2$.

Note that $X > 0$ (since $S \neq \{1, \ldots J'\}$) and recall our assumption that every $x_i > \epsilon$. For each $i = 1, 2, \ldots k$, by (4.14) and since $\epsilon < X + \epsilon < 1$,

$$(X + x_i)^{r'} > (X + \epsilon)^{r'} > 4.2^{J'}.X^{r'} \tag{4.16}$$

Since $\sum_{i=1}^{k}(P - s_i) = (k-1).P$ at least one of the $P - s_i$'s has modulus not less than $\frac{(k-1)|P|}{k}$, so at least $\frac{|P|}{2}$. Squaring gives $(P - s_i)^2 \geq \frac{P^2}{4}$. This identifies a positive term above larger than

$$\frac{P^2.4.2^{J'}.X^{r'}}{4} = 2^{J'}P^2X^{r'} \tag{4.17}$$

outweighing all of the negative terms and proving the claim.

Hence the lemma is proved.

■

Notation In the following section we write down knowledge bases without using dezeroed notation. However, dezeroed notation returns in Section 4.3.

4.2 Defining $Maximin^L$

Definition The inference process $Maximin^L$ is defined analogously to $Minimax^L$. Let $\tilde{\vec{x}}$ = the unique vector $\tilde{\vec{x}}$ which is a permutation of $\vec{x}$ for which $\tilde{x}_1 \leq \tilde{x}_2 \leq \ldots \leq \tilde{x}_J$. It should be clear from the context whether $\tilde{}$ is being used in this sense, or in its previous usage from Chapter 2, referring to $Minimax$. We define the *maximin* ordering on vectors in $\mathbb{R}^J$ by: $\vec{x}$ is before $\vec{y}$ in the *maximin* ordering iff $\tilde{\vec{x}} > \tilde{\vec{y}}$ lexicographically. We shall say that $\vec{x}$ is *maximin*-better in that case.

For any consistent knowledge base K,

$$\begin{aligned} Maximin^L(K) &= \text{the unique } \vec{x} \in V^L(K) \text{ for which } \tilde{\vec{x}} \text{ is maximal lexicographically} \\ &= \text{that } \vec{x} \in V^L(K) \text{ which is } maximin\text{-best.} \end{aligned} \quad (4.18)$$

In other words, $\widetilde{Maximin^L(K)}_1$ is maximised and subject to that condition $\widetilde{Maximin^L(K)}_2$ is maximised and so on.

When we calculate $Maximin^L$ we may use the abbreviation $Mxmn^L(*)$.

Range of notation The terms that refer to $Maximin^L$ explicitly are defined here for use throughout this thesis, namely the *maximin* ordering, *maximin*-better, *maximin*-best.

In Section 4.3 we shall prove the central theorem of this chapter:

Theorem 64 *Given $K \in CL$, there exists N such that for all $r > N$, $Every_r^L(K)$ is well-defined and $lim_{r\to\infty} Every_r^L(K) = Mxmn^L(K)$.*

Thus, assuming the philosophical basis for the $Every_r^L$ "inference processes", we find that they are well defined in the limit as r tends to infinity and we can justify using the inference process $Maximin^L$. We now prove that $Maximin^L$ is well-defined, by a similar method to that by which we have shown in Chapter 2 that $Minimax^L$ is well-defined.

Definition In this chapter, and elsewhere if the context refers to $Maximin^L$, $m_1, m_2 \ldots m_J$ are functions of $K \in CL$ given by:

$$m_1 = max\{min_{1\leq i\leq J} x_i \mid \vec{x} \in V^L(K)\} \quad (4.19)$$

and then

$m_{k+1} = max\{min\ x_i\,|$ there exist $i_1, i_2, \dots i_k$ distinct from i and from each other and s.t. $x_{i_1}, x_{i_2}, \dots x_{i_k}$ are equal to $m_1, m_2, \dots m_k$ respectively where $\vec{x} \in V^L(K)\}$. Then we can see that $Mxmn^L(K)$ is a vector in $\vec{V_L}(K)$ which is a permutation of $(m_1, m_2, \dots m_J)$.

Lemma 65 *For every $k = 1, 2, \dots J$, $m_1 \leq m_2 \dots \leq m_k$ and there exists $\vec{x} \in V^L(K)$ s.t. $\tilde{x}_1 = m_1, \dots \tilde{x}_k = m_k$. In other words, there are distinct co-ordinates $i_1, \dots i_k$ such that $x_{i_p} = m_p$ for each $p = 1, \dots k$ and no other value in $\vec{x}$ is less than m_k.*

Theorem 66 *There exists a bijection $\sigma : \{1, 2, \dots J\} \to \{1, 2, \dots J\}$ such that for each $k = 1, 2, \dots J$ and all $\vec{x} \in V^L(K)$ s.t. $\tilde{x}_1 = m_1, \dots \tilde{x}_k = m_k$ then $x_{\sigma(i)} = m_i$ for all $i = 1, 2, \dots k$.*

Definition If, given K, the identity permutation can fulfil the role of σ above, K is said to *admit the identity permutation w.r.t. Theorem* 66.

Corollary 67 *The inference process $Maximin^L$ is well-defined.*

Proofs We can prove Lemma 65, Theorem 66 and Corollary 67 by reversing some inequalities in the proofs of Lemma 27, Theorem 28 and Corollary 29 respectively.

■

4.3 Showing that $lim_{n\to\infty}Every_n^L(K) = Maximin^L(K)$

Theorem 64 *Given $K \in CL$, there exists N such that for all $r > N$, $Every_r^L(K)$ is well-defined and $lim_{r\to\infty}Every_r^L(K) = Mxmn^L(K)$.*

Notation Throughout this section, we assume that $K \in CL$ is a fixed knowledge base and we assume w.l.o.g. that we write it in dezeroed form, just as in Section 4.1. We enumerate the atoms of L, $\alpha_1, \ldots \alpha_J$ in such a way that

$$I^L(K) = \{i \text{ s.t. } J' < i \leq J\} \tag{4.20}$$

w.l.o.g. since $Every_n^L$ and $Mxmn^L$ are defined in a manner symmetrical w.r.t. the labelling of the atoms. Also, w.l.o.g., K includes constraints of the form $x_i = 0$ for each i s.t. $J' < i \leq J$ and every other constraint of K only refers to $x_1, x_2, \ldots x_{J'}$. We write solutions of K as points in $\mathbb{D}^{J'}$, ignoring the constant zeros.

Recall that

$$Every_r^L(K) = \text{the unique } \vec{x} \in V^L(K) \text{ such that } All(\vec{x}, r) \text{ is maximal} \tag{4.21}$$

if such an $\vec{x}$ is unique, otherwise it is not defined. We further assume w.l.o.g. that, not using dezeroed notation,

$$Mxmn^L(K) = (m_{J-J'+1}, m_{J-J'+2}, \ldots m_J, m_1, m_2, \ldots m_{J-J'}) \tag{4.22}$$

and we can do this by an appropriate enumeration of the atoms of L, because the smallest $J - J'$ values of (4.22) must all be 0 and occur, by (4.20), at the last $J - J'$ co-ordinates. Following the definition of $Maximin^L$, of those solutions of K for which the smallest $J - J'$ values are all zero (i.e. all solutions of K), $m_{J-J'+1}$ is the largest possible $min_{i=1\ldots J'}x_i$ etc.. We see that $(m_{J-J'+1}, \ldots m_J)$ is the *maximin*-best $(x_1, \ldots x_{J'})$ for which $\vec{x} \in V^L(K)$. In dezeroed notation, we rewrite (4.22) thus:

$$Mxmn^L(K) = (M_1, M_2, \ldots M_{J'}) \tag{4.23}$$

where $0 < M_1 < M_2 < \ldots < M_{J'}$ (since $Maximin^L$ satisfies Open-mindedness (Theorem 78)).

During the proof below other knowledge bases $K^{(r)}$ etc. are introduced for which $I^L(K) \subseteq I^L(K^{(r)})$ etc., but the notation remains dezeroed in the sense that we ignore the co-ordinates in $I^L(K)$, rather than all of those in $I^L(K^{(r)})$. From now on in this section, "minimum" and the function min refer to $min_{i=1\ldots J'}$. The second smallest value, etc. is also calculated after ignoring the co-ordinates in $I^L(K)$.

M_1 is the largest possible minimum of a solution of K and, of those $\vec{x} \in V^L(K)$ for which $min(\vec{x}) = M_1$, M_2 is the largest possible second smallest value, and so on. By Theorem 65, for each $1 \leq k \leq J'$, if the smallest k values of $\vec{x}$ are $M_1, \ldots M_k$ and $\vec{x} \in V^L(K)$, then $x_i = M_i$ for all $k = 1, \ldots J'$. In other words, by our assumptions, K "after ignoring co-ordinates in $I^L(K)$, admits the identity permutation w.r.t. Theorem 66".

Proof of Theorem 64 To start with, even though points minimising $Nall(\vec{x}, r)$ might not be unique they do exist because $Nall(\vec{x}, r)$ is continuous and $V^L(K)$ is closed. Define

$$\widehat{Every_r}^L(K) = \{\vec{x} \in V^L(K) | \text{ for all } \vec{y} \in V^L(K), Nall(\vec{x}, r) \leq Nall(\vec{y}, r)\} \tag{4.24}$$

which is also closed and non-empty so
$min(\widehat{Every_r}^L(K)) = min\{min(\vec{x}) \,|\, \vec{x} \in \widehat{Every_r}^L(K)\}$ is well-defined.

Lemma 68 *As $r \to \infty$, $min(\widehat{Every_r}(K)) \to M_1$*

Proof Given positive ϵ less than $\frac{1}{2}$, choose $N > \frac{\log J'}{\log(\frac{1}{1-\epsilon})}$ and, just as for (2.13) in the proof of Lemma 30, we see that

$$x^r > J'(x - \epsilon)^r \tag{4.25}$$

for all $x \in (\epsilon, 1)$ and each $r > N$. Now suppose for contradiction that there exists $\vec{x} \in V^L(K)$ such that $Nall(\vec{x}, r)$ is minimal but $min(\vec{x}) = x_{i*} < M_1 - \epsilon$. Let $\vec{y} \in V^L(K)$ be such that $min(\vec{y}) = M_1 \leq \frac{1}{2}$. Also

$$\epsilon < \frac{1}{2} \implies 1 - x_{i*} - \epsilon > 1 - M_1 > 0 \tag{4.26}$$

Then, looking at the definition of $Nall(\vec{y}, r)$ as a probability,

$$\begin{aligned} Nall(\vec{y}, r) &\leq (1 - y_1)^r + (1 - y_2)^r + \ldots + (1 - y_{J'})^r \\ &\leq J'(1 - m_1)^r \leq J'(1 - x_{i*} - \epsilon)^r \\ &< (1 - x_{i*})^r \text{ by (4.25)} \\ &\leq Nall(\vec{x}, r) \end{aligned} \tag{4.27}$$

contradicting minimality of $Nall(\vec{x}, r)$. Hence every minimum of a vector in $\widehat{Every_r}^L(K)$ is at least $M_1 - \epsilon$ so we have proved the lemma.

■

Proof of Theorem 64 Recall that $M_1 > 0$. By Lemma 68 for some $N_1 \in \mathbb{N}$ and every $r > N_1$, every $\vec{x}$ for which $Nall(\vec{x}, r)$ is minimal satisfies the condition that $min(\vec{x}) > M_1/2$.

Using Lemma 63 let N_2 be such that for every $r > N_2$ $Nall(\vec{x}, r)$ is convex in $\mathbb{D}^{J'}_{>M_1/2}$, which we define thus:

$$\mathbb{D}^{J'}_{>M_1/2} = \{\vec{x} \in \mathbb{D}^{J'} \text{ s.t. } x_i > M_1/2 \text{ for } i = 1, \ldots J'\} \tag{4.28}$$

so if there exist distinct $\vec{X}, \vec{Y}$ in $V^L(K)$ where the value of $Nall(\vec{x}, r)$ is minimal for some $r > max(N_1, N_2)$ they are both in $\mathbb{D}^{J}_{>M_1/2}$. By strict convexity of $Nall$, and of $V^L(K)$, $\frac{1}{2}(\vec{X} + \vec{Y})$ has a smaller value of $Nall$, which is a contradiction. Thus, if $r > max(N_1, N_2)$, $Every_r(K)$ is well-defined.

Notation In this section, we let $x_i^{(r)}$ denote $Every_r(K)_i$ and let $min^{(r)} = min_{i=1}^{J'} x_i^{(r)}$.

Given that we have shown that for large enough r, $Every_r(K)$ is well-defined, we can now restate the remaining content of Theorem 64 as:

$$\text{For each } i = 1, 2, \ldots J',\ x_i^{(r)} \to M_i \text{ as } r \to \infty$$

which we prove by strong induction on i.

Base Case Lemma 68 implies that $min(Every_r(K)) \to M_1$ as $r \to \infty$. In a similar way to the Base Case of the proof of Theorem 26, using selected subsequences and compactness, we can show that $x_1^{(r)} \to M_1$ as $r \to \infty$.

Inductive Step Suppose (I.H.) that $x_1^{(r)} \to M_1, \ldots x_k^{(r)} \to M_k$ as $r \to \infty$, for some k such that $1 \leq k < J'$.

Definition For each integer $k = 1, \ldots J'$, define the function g_k as follows, where the input $K' \in CL$ is such that $I^L(K) \subseteq I^L(K')$:

$$g_k = max\{min_{k \leq i \leq J'} x_i \mid \vec{x} \in V^L(K')\} \tag{4.29}$$

Lemma 69 *The functions g_k are uniformly continuous in the Blaschke topology. In fact, for each $k = 1, \ldots J'$ and all $K_1, K_2 \in CL$ for which $I^L(K) \subseteq I^L(K_1)$ and $I^L(K) \subseteq I^L(K_2)$, $|\, g_k(K_1) - g_k(K_2)| \leq \Delta(K_1, K_2)$.*

The proof is similar to that of Lemma 31 with inequalities reversed.

■

Proof of Theorem 64 (Inductive Step) continued If ϵ is given s.t. $\epsilon \in (0, M_1/2)$, let

$$K^{(r)} = K + \left\{x_1 = x_1^{(r)}, x_2 = x_2^{(r)}, \ldots x_k = x_k^{(r)}\right\} \tag{4.30}$$

and

$$K^{(\infty)} = K + \{x_1 = M_1, \ldots x_k = M_k\} \tag{4.31}$$

We note that $K^{(r)}$, $K^{(\infty)}$ include the constraints $x_i = 0$ for $J' < i \leq J$ and we write solutions of these knowledge bases as points in $\mathbb{D}^{J'}$. Since $K^{(r)} \to K^{(\infty)}$ as $r \to \infty$, (by Theorem 25) there exists $N_3 \in \mathbb{N}$ such that for $r > N_3$, $\Delta(K^{(r)}, K^{(\infty)}) < \epsilon/2$ so

$$|\, g_k(K^{(r)}) - g_k(K^{(\infty)})| = |\, g_k(K^{(r)}) - M_{k+1}| < \epsilon/2 \tag{4.32}$$

by Lemma 69. Now there exists $N_4 \in \mathbb{N}$ s.t. if $r > N_4$ and $i \leq k$ then $x_i^{(r)} > M_1/2$ since the $x_i^{(r)}$ are tending to limits $M_i \geq M_1 > 0$ and $g_k(K^{(r)}) > M_1/2$. Suppose that

$$r > \hat{N} = max\left(N_1, N_2, N_3, N_4, \frac{\log 2J'}{\log(\frac{2}{2-\epsilon})}\right) \tag{4.33}$$

and let $min^{(k,r)}$ denote $min_{i=k+1,\ldots J'} x_i^{(r)}$. Note that $Every_r$ is well-defined on $K^{(r)}$ when it is well-defined on K and in this case $Every_r(K^{(r)}) = Every_r(K)$.

We now assume for contradiction that for some fixed $r = r_0 > \hat{N}$, $|g_k(K^{(r_0)}) - min^{(k,r_0)}| > \epsilon/2$ so that $min^{(k,r_0)} < g_k(K^{(r_0)}) - \epsilon/2$.

Notation Let $\vec{X} = \vec{x}^{(r_0)}$ and let $i = i_0$ be fixed such that $i_0 > k$ and $X_{i_0} = min^{(k,r_0)}$. Let $\vec{Y} \in V(K^{(r_0)})$ s.t. $min_{i=k+1\ldots J'} Y_i = g_k(K^{(r_0)})$.

Definition Define the function $Somefail(\vec{x})$ = *Probability that in r_0 trials of the p.b. function $\vec{x}$, at least one of the atoms $\alpha_1, \ldots \alpha_k$ fails to occur.* We see that $Somefail(\vec{X}) = Somefail(\vec{Y})$, say $= Sf$. Now

$$r_0 > \frac{\log 2J'}{\log(\frac{2}{2-\epsilon})} \tag{4.34}$$

ensuring that if $x \in [\epsilon/2, 1]$,

$$x^{r_0} > 2J'(x - \epsilon/2)^{r_0} \tag{4.35}$$

in a similar way to (4.25). Then using the Inclusion-Exclusion Principle leads to

$$\begin{aligned} Nall(\vec{Y}, r_0) &\leq Sf + \text{ the probability that in } r_0 \text{ trials outcome } k+1 \text{ or ..., or } J' \\ &\quad\ \text{fails} \\ &\leq Sf + \sum_{w=k+1}^{J'} (1 - Y_w)^{r_0} \leq Sf + (J' - k)(1 - g_k(K^{(r_0)}))^{r_0} \\ &< Sf + \frac{1}{2}(1 - g_k(K^{(r_0)}) + \epsilon/2)^{r_0} \end{aligned} \tag{4.36}$$

using (4.35). Now for all $\vec{x} \in V^L(K)$, let $F(\vec{x})$ be the probability that at least one of the first k atoms or the i_0'th fail to occur in r_0 trials. By the Inclusion-Exclusion Principle,

$$\begin{aligned} &F(\vec{X}) \geq Sf + \text{ the probability that the } i_0\text{'th outcome fails} \\ &\quad \text{minus the probability that the } i_0\text{'th and one of the first } k \text{ outcomes fail} \\ \geq\ & Sf + (1 - X_{i_0})^{r_0} - (1 - X_1 - X_{i_0})^{r_0} - (1 - X_2 - X_{i_0})^{r_0} \ldots - (1 - X_k - X_{i_0})^{r_0} \\ \geq\ & Sf + (1 - X_{i_0})^{r_0} - J'(1 - m_1/2 - X_{i_0})^{r_0} \\ \geq\ & Sf + \frac{1}{2}(1 - X_{i_0})^{r_0} \end{aligned} \tag{4.37}$$

since $X_1, \ldots X_k$ are all not less than $M_1/2$ and, using $\frac{\epsilon}{2} < \frac{M_1}{2}$ and (4.34),

$$r_0 > \frac{\log 2J'}{\log(\frac{2}{2-M_1})} \tag{4.38}$$

We combine this with (4.36) to give

$$\begin{aligned} Nall(\vec{Y}, r_0) &< Sf + \frac{1}{2}(1 - g_k(K^{(r_0)}) + \epsilon/2)^{r_0} \\ &\leq Sf + \frac{1}{2}(1 - X_i)^{r_0} \leq F(\vec{X}) \leq Nall(\vec{X}, r_0) \end{aligned} \tag{4.39}$$

(by inspecting the definitions), contradicting the definition of $\vec{X}$. Hence the assumption that $min^{(k,r_0)} < g_k(K^{(r_0)}) - \epsilon/2$ is false so $|\, min^{(k,r)} - g_k(K^{(r)})| < \epsilon/2$ for all $r > \hat{N}$.

Finally, using (4.32) gives $|min^{(k,r)} - M_{k+1}| < \epsilon$, so $min^{(k,r)} \to M_{k+1}$ as $r \to \infty$. Then using compactness and subsequences again leads to $x^{(r)}_{k+1} \to M_{k+1}$ so we have proved the Inductive Step and Theorem 64.

■

Chapter 5

The properties of $Maximin^L$ compared with $Minimax^L$

5.1 Comparing $Maximin^L$ with $Minimax^L$ and ME^L

We carry out a similar comparison to that of Chapter 3 and test $Maximin^L$ against the Par-Ven Properties—those properties of inference processes used in [ParVen1] to uniquely characterise *Maximum Entropy*. Due to the dual nature of $Minimax^L$ and $Maximin^L$ they share many properties but there are also some significant differences in their behaviour.

5.1.1 Equivalence

Theorem 70 *$Maximin^L$ satisfies Equivalence.*

Proof The definition of $Maximin^L(K)$ is given in terms of $V^L(K)$ so Equivalence is satisfied by $Maximin^L$.

$\blacksquare$

5.1.2 Atomic Renaming

Theorem 71 *$Maximin^L$ satisfies Atomic Renaming.*

Proof The definition of $Maximin^L$ is symmetrical w.r.t. permutations of the atoms so Atomic Renaming holds for $Maximin^L$.

■

5.1.3 Obstinacy

Theorem 72 *$Maximin^L$ satisfies Obstinacy.*

Proof For any consistent knowledge base K, $Mxmn^L(K)$ is the optimal solution of K w.r.t. a fixed partial ordering, namely the *maximin* ordering. Hence, by Theorem 9, $Maximin^L$ satisfies Obstinacy and we have proved the theorem.

■

Remark Note that $Maximin^L$ cannot be expressed as minimising a real-valued function over $V^L(K)$ for essentially the same reason as $Minimax^L$:
see Theorem 38.

5.1.4 Language Invariance

Theorem 73 *$Maximin^L$ is Language Invariant.*

Proof The following lemma will be useful in proving that $Maximin^L$ has certain properties:

Lemma 74 *Let $C \subset \{1, \dots J\}$ and let $\sim$ be an equivalence relation on $\{1, \dots J\} \setminus C$ such that the equivalence classes are all of equal size. Let $i_1, \dots i_q$ be representatives of the q distinct equivalence classes. For each $\vec{x} \in \mathbb{D}^J$ s.t. $x_i = x_j$ for all i, j s.t. $i \sim j$, let $Simp(\vec{x}) = (x_{i_1}, \dots x_{i_q})$. Then if $\vec{y} \in \mathbb{R}^J$ is such that $y_c = x_c$ for all $c \in C$, comparing $\vec{x}$ and $\vec{y}$ in the maximin ordering is equivalent to comparing $Simp(\vec{x})$ and $Simp(\vec{y})$ in the maximin ordering.*

Corollary 75 *Suppose $\vec{x}, \vec{y}$ are vectors in $\mathbb{R}^J$ such that $x_i = y_i$ for all $i \in C$, for some $C \subset \{1, \dots J\}$. W.l.o.g. let $C = \{1, \dots k\}$. We can do this by the symmetry of the maximin ordering w.r.t. permuting the co-ordinates. Then comparing $\vec{x}$ and $\vec{y}$*

in the maximin ordering is equivalent to comparing $(x_{k+1}, \dots x_J)$ *and* $(y_{k+1}, \dots y_J)$. *If the same* C *is a subset of co-ordinates which are constant w.r.t. a knowledge base* $K \in CL$, *then* $Maximin^L(K) =$ *that* $\vec{x} \in V^L(K)$ *for which* $(x_{k+1}, \dots x_J)$ *is maximin-best.*

Proofs of Lemma 74, Corollary 75 These are essentially the same as the proofs of Lemma 42 and Corollary 43 respectively, with some inequalities reversed.

■

Proof of Theorem 73 If $Maximin^L$ agrees with $Maximin^{L'}$ in the cases for which $L' = L + \{p'\}$, for some $p' \notin L$, then it satisfies Language Invariance since we can modify any language into any other by a step-by-step process, adding or removing one p.v. at each step.

We assume for a contradiction that $At^L = \{\alpha_1, \dots \alpha_J\}$ and that there exists $K \in CL$ s.t. $Mxmn^{L'}(K)(\alpha_i) \neq Mxmn^L(K)(\alpha_i)$ for some $i = 1, \dots J$, where L' is a language of the form $L' = L + \{p'\}$ and $p' \notin L$. For each $\alpha \in At^L$, there exist 2 atoms of L', β, such that $\beta \models \alpha$.

We label the atoms of L' thus, up to logical equivalence:

$$At^{L'} = \{\beta_{(1,+)}, \beta_{(1,-)}, \beta_{(2,+)}, \beta_{(2,-)} \dots \beta_{(J,+)}, \beta_{(J,-)}\} \tag{5.1}$$

where $\beta_{(i,+)} = \alpha_i \wedge p'$ and $\beta_{(i,-)} = \alpha_i \wedge \neg p'$. By Atomic Renaming, as K does not mention p', $Mxmn^{L'}(K) = \vec{W} \in \mathbb{D}^{2J}$ gives the same belief value to $\beta_{(i,+)}$ and $\beta_{(i,-)}$ for each i. Let $\vec{y} = Mxmn^L(K)$ and define $w_i = Mxmn^{L'}(K)(\alpha_i)$ for $i = 1, \dots J$, so that $\vec{w}$, $\vec{y} \in V^L(K)$. By uniqueness of $Maximin^L(K)$ and our assumption, $\vec{y}$ is $maximin$-better than $\vec{w}$.

Now we let $Simp$ be as in Lemma 74, where $\sim$ is the equivalence relation on the co-ordinates $i, \pm$ such that $\sim$ has J equivalence classes of the form $\{(i,-), (i,+)\}$ for $i = 1, \dots J$. C is empty and 2 is the size of all the equivalence classes. We know that $\vec{y}/2 = Simp(\vec{Y})$ is $maximin$-better than $\vec{w}/2 = Simp(\vec{W})$. Also $\vec{Y} \in V^{L'}(K)$ since it is a way of splitting up the belief values given by $\vec{y}$ between the atoms of L'.

By Lemma 74, $\vec{Y}$ is *maximin*-better than $\vec{W} \in V^{L'}(K)$, contradicting the fact that $\vec{W} = Mxmn^{L'}(K)$. Hence $Maximin^L$ is Language Invariant and we have proved the theorem.

■

From this point on we will usually refer to $Maximin$ without mentioning the overlying language.

5.1.5 Continuity

Definition The function m_1 is defined on all consistent knowledge bases: if $K \in CL$ then $m_1(K) = max_{\vec{x} \in V^L(K)} min_{i=1\ldots J} x_i$.

Lemma 76 *The function m_1 is uniformly continuous in the Blaschke topology. In fact, for every $K_1, K_2 \in CL$, $|\, m_1(K_1) - m_1(K_2)| \leq \Delta(K_1, K_2)$.*

The proof is just as for Lemma 31 with inequalities reversed.

■

By Lemma 76, the function m_1 (as defined by (4.19)) is uniformly continuous in the Blaschke topology. However, like $Minimax$, $Maximin$ fails to satisfy Continuity.

Theorem 77 *$Maximin$ is not continuous.*

Proof We define $K_\epsilon, Sol_\epsilon, \vec{s}(\epsilon, \tau)$ in exactly the same way as the proof of Theorem 40 and see that $K_\epsilon \to K_0$ as $\epsilon \searrow 0$.

For $\epsilon > 0$, $min(\vec{s}(\epsilon, \tau))$ = either $\epsilon\tau$ or $\frac{1}{2} - \tau$. With ϵ fixed, the minimum of these two values is maximal when they are equal, since $\frac{1}{2} - \tau$ decreases as $\epsilon\tau$ increases.

Hence for the *maximin*-best solution of K_ϵ, we require that $\tau = \frac{1}{2(1+\epsilon)}$ and

$$Mxmn(K_\epsilon) = \left(\frac{\epsilon}{2(1+\epsilon)}, \frac{1}{2(1+\epsilon)}, \frac{\epsilon}{2(1+\epsilon)}, \frac{1}{2(1+\epsilon)}\right) \tag{5.2}$$

which tends to $(0, \frac{1}{2}, 0, \frac{1}{2})$ as $\epsilon \searrow 0$.

However, for $\epsilon = 0$, $Mxmn(K_0) = (0, \frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ so $Maximin$ is not continuous and we have proved the theorem. ■

5.1.6 Open-mindedness

Theorem 78 *Maximin satisfies Open-mindedness.*

Proof Let $|I^L(K)| = num$. We use $m_1, \dots m_J$ as defined close to (4.19). Then $m_1 = m_2 = \dots = m_{num} = 0$. By [[Par],page 74], there exists $\vec{x} \in V^L(K)$ such that $x_i > 0$ for each $i \notin I^L(K)$. Hence $m_{num+1} > 0$, by definition. Since the m_i's are an increasing sequence (Lemma 65) and $Mxmn(K)$ is a permutation of $(m_1, \dots m_J)$, $Mxmn(K)_i > 0$ for all $i \notin I^L(K)$ and we have proved that $Maximin$ satisfies Open-mindedness.

■

5.1.7 Independence

Theorem 79 *Maximin does not satisfy Independence.*

Remark Since $Maximin$ is the dual of $Minimax$, this result should not be surprising. However, the largest distance between $Maximin$ and the Independent solution is significantly smaller than in the case of $Minimax$!

We contain the proof of Theorem 79 within an investigation of how close $Maximin$ comes to satisfying Independence which follows the same methods as Subsection 3.1.7.

How close does $Maximin$ come to giving the Independent solution?

We will look later at the knowledge bases used in the definition of Independence.

For simplicity consider first a knowledge base $K_{b,c}$ of the form:

$$K_{b,c} = \{x_1 + x_2 = b,\ x_1 + x_3 = c\} \tag{5.3}$$

where $x_i = Bel(\alpha_i)$ for $i = 1, 2, 3, 4$.

Notation Let the *Independent solution* of $K_{b,c}$ be denoted by

$$Ind(b, c) = ME(K_{b,c}) = (bc, b(1-c), (1-b)c, (1-b)(1-c)) \tag{5.4}$$

for all $b \in [0, 1]$ and $c \in [0, 1]$. Also we denote $Maximin(K_{b,c})$ by $Mxmn(b, c)$.

Recall from Subsection 3.1.7 that $K_{b,c}$ is consistent iff $b \in [0,1]$ and $c \in [0,1]$.

Just as for $Minimax$ we assume w.l.o.g. that $0 \le b \le c \le \frac{1}{2}$.

Notation We denote $V^L(K_{b,c})$ by $Sol(b,c)$. Recall that

$$Sol(b,c) = \{(\tau, b-\tau, c-\tau, 1-b-c+\tau) \mid 0 \le \tau \le b\} \tag{5.5}$$

The minimum of a vector of the above form is either τ or $b-\tau$ since $b \le c$ and $b+c \le 1$. The minimum is maximised when $\tau = b/2$ so

$$\begin{aligned} Mxmn(b,c) = Mxmn(K_{b,c}) &= \frac{1}{2}(b, b, 2c-b, 2-b-2c) \\ \text{Also, } |Mxmn(b,c) - Ind(b,c)| &= |2bc - b| \end{aligned} \tag{5.6}$$

This is due to (3.20); the difference between the values at the first co-ordinate is $|bc - \frac{b}{2}|$.

Hence, assuming the convention that $0 \le b \le c \le \frac{1}{2}$, $Mxmn(b,c) = Ind(b,c)$ iff $c = \frac{1}{2}$ or $b = 0$.

Theorem 80 *For $b \in [0,1]$, $c \in [0,1]$, $Mxmn(b,c) = Ind(b,c)$ iff either b or c equal either 0, $\frac{1}{2}$ or 1.*

Proof This is similar to the proof of Theorem 45.

■

Remark So far, apart from the fact that we have not needed to split cases as in (3.18), $Maximin$ has performed just as well as $Minimax$. However in the following theorem $Maximin$ shows a significant improvement on $Minimax$.

Theorem 81 *If $b \in [0,1]$, $c \in [0,1]$, $|Mxmn(b,c) - Ind(b,c)|$ takes its maximal value, $\frac{1}{8}$, iff $(b,c) = \left(\frac{1}{4}, \frac{1}{4}\right), \left(\frac{1}{4}, \frac{3}{4}\right), \left(\frac{3}{4}, \frac{1}{4}\right)$ or $\left(\frac{3}{4}, \frac{3}{4}\right)$.*

Proof We again assume that $0 \le b \le c \le 1/2$.

The largest value of $|Mxmn(b,c) - Ind(b,c)|$ occurs iff $|2bc - b| = b(1-2c)$ is maximised. Let the triangular region R_3 of the b,c plane be given by

$$R_3 = \left\{(b,c) \text{ s.t. } 0 \le b \le c \le \frac{1}{2}\right\} \tag{5.7}$$

Consider a point in R_3. If it is not on the edge given by $b = c$ then, if we increase b and keep c constant, $b(1-2c)$ increases. All of the "worst" points lie on the edge of R_3 given by

$$b = c,\ 0 \leq b \leq 1/2 \tag{5.8}$$

Given these conditions, $b - 2b^2$ is maximal at $b = \frac{1}{4}$ so $b = c = \frac{1}{4}$ gives the biggest value of $|Ind(b,c) - Mxmn(b,c)|$.

Discarding the assumption that $0 \leq b \leq c \leq 1/2$, i.e. as we let b, c take values throughout $[0,1]$, the knowledge bases $K_{b,c}$ producing the largest distances between $Ind(b,c)$ and $Mxmn(b,c)$ are given by $(b,c) = (\frac{1}{4},\frac{1}{4}), (\frac{1}{4},\frac{3}{4}), (\frac{3}{4},\frac{1}{4})$ and $(\frac{3}{4},\frac{3}{4})$. In these cases $|Ind(b,c) - Mxmn(b,c)| = 2(\frac{1}{8} - \frac{1}{16}) = \frac{1}{8}$ so we have proved the theorem.

■

The knowledge bases used in the definition of Independence

Now, for each $a, b, c \in [0,1]$ s.t. $a > 0$ we define

$$K_{a,b,c} = \{Bel(p_1) = a, Bel(p_2|p_1) = b, Bel(p_3|p_1) = c\} \tag{5.9}$$

as in [[Par], page 101]. The Independent solution is defined by $Ind(a,b,c) = ME(K_{a,b,c})$ as in Subsection 3.1.7.

In the following theorem, $Mxmn(a,b,c)$ denotes $Maximin(K_{a,b,c})$.

Theorem 82 *If $K_{a,b,c}$ is consistent, $|Mxmn(a,b,c) - Ind(a,b,c)|$ takes its maximal value, $\frac{1}{8}$, when $a = 1$ and $(b,c) = \left(\frac{1}{4},\frac{1}{4}\right), \left(\frac{1}{4},\frac{3}{4}\right), \left(\frac{3}{4},\frac{1}{4}\right)$ or $\left(\frac{3}{4},\frac{3}{4}\right)$.*

Proof We can proceed as in the proof of Theorem 47, apart from a few modifications. As far as the Claim at 3.33 we need only replace Mmx by $Mxmn$.

Claim

$$Mxmn(a,b,c)_{\mathbf{4}} = a.Mxmn(b,c) \tag{5.10}$$

Proof of claim We use Corollary 75 instead of Corollary 43, replacing Mmx by $Mxmn$ etc. in the proof of the claim at (3.33).

■

Theorem 82 now follows similarly to Theorem 47. ■

Proof of Theorem 79 This follows from Theorem 82.

■

5.1.8 Relativisation

Theorem 83 *Maximin satisfies Relativisation.*

Proof We fix a language L and a sentence ϕ as in the definition of Relativisation, in section 1.4.8. Let K_1 be a knowledge base given by

$$K_1 = \{Bel(\phi) = c\} \cup \left\{ \sum_{j=1}^{s} a_{ji} Bel(\theta_j|\phi) = b_i \,|\, i = 1, \dots m \right\} \tag{5.11}$$

for some constants c, s, a_{ji}, b_i and sentences θ_i.

We choose, w.l.o.g. (by Theorem 71), to enumerate the atoms of L, α_i, such that

$$\{\alpha_1, \alpha_2, \dots \alpha_k\} = \{\alpha_i \in At^L \text{ s.t. } \alpha_i \models \phi\} \tag{5.12}$$

Now suppose that

$$K_2 = K_1 + \left\{ \sum_{j=1}^{t} c_{ji} Bel(\theta_j|\neg\phi) = d_i \,|\, i = 1, \dots p \right\} \tag{5.13}$$

When we write $x_i = Bel(\alpha_i)$ for each $i = 1, \dots J$, K_1 takes the form

$$K_1 = \left\{ \sum_{i=1}^{k} x_i = c, \sum_{i=k+1}^{J} x_i = 1 - c \right\} \cup \left\{ \sum_{i=1}^{k} f_{ij} x_i = g_j \,|\, j = 1, \dots m \right\} \tag{5.14}$$

for some constants c, f_{ij}, g_j. Similarly,

$$K_2 = K_1 + \left\{ \sum_{i=k+1}^{J} f_{ij} x_i = g_j \,|\, j = m+1, \dots m+p \right\} \tag{5.15}$$

Let $Mxmn(K_1) = \vec{X}$ and $Mxmn(K_2) = \vec{Y}$. For any $\vec{x} \in \mathbb{D}^J$, let $\vec{x}_{\mathbf{k}}$ denote $(x_1, \dots x_k)$. Then we suppose for contradiction that

$$\vec{X}_{\mathbf{k}} \neq \vec{Y}_{\mathbf{k}} \tag{5.16}$$

Define $\overline{XtoY} = (X_1, \dots X_k, Y_{k+1}, \dots Y_J)$ and $\overline{YtoX} = (Y_1, \dots Y_k, X_{k+1}, \dots X_J)$. We see that $\overline{XtoY} \in V^L(K_2)$ and $\overline{YtoX} \in V^L(K_1)$. This is because the constraints of K_1

only mention $x_1, \dots x_k$, apart from $\sum_{k+1}^{J} x_i = 1 - c$ and the additional constraints in K_2 only mention $x_{k+1}, \dots x_J$.

Since $\vec{Y} \neq \overline{XtoY} \in V^L(K_2)$, $\vec{Y}$ is *maximin*-better than $\overline{XtoY}$. Using Corollary 75, with $C = \{k+1, \dots J\}$, we deduce that $\vec{Y_{\mathbf{k}}}$ is *maximin*-better than $\vec{X_{\mathbf{k}}}$.

Since $\vec{X} \neq \overline{YtoX} \in V^L(K_1)$, $\vec{X}$ is *maximin*-better than $\overline{YtoX}$. Using Corollary 75, with $C = \{k+1, \dots J\}$, we deduce that $\vec{X_{\mathbf{k}}}$ is *maximin*-better than $\vec{Y_{\mathbf{k}}}$, contradicting the fact that $\vec{Y_{\mathbf{k}}}$ is *maximin*-better than $\vec{X_{\mathbf{k}}}$.

Hence our assumption (5.16) is false so $Maximin$ satisfies Relativisation and we have proved the theorem.

■

5.1.9 Irrelevant Information

Theorem 84 *Maximin satisfies Irrelevant Information.*

The fact that Irrelevant Information is a property of $Maximin$ is the most striking advantage $Maximin$ has over $Minimax$, when we compare the properties of these inference processes with the list of desiderata which uniquely specify $Maximum$ $Entropy$. $Maximum$ $Entropy$ is not as lonely in satisfying Irrelevant Information and Atomic Renaming as was previously thought!

Notation The following notation applies to this subsection.

Suppose that $K_1 \in CL_1$, $K_2 \in CL_2$, $L_1 \cap L_2 = \emptyset$ and $L_1 \cup L_2 = L$. Let J denote the number of atoms α_i in At^{L_1} and let Q denote the number of atoms β_j in At^{L_2}. Note that

$$At^L = \{\alpha_i \wedge \beta_j \text{ s.t. } \alpha_i \in At^{L_1}, \beta_j \in At^{L_2}\} \tag{5.17}$$

up to logical equivalence.

Now we express each solution of $K_1 + K_2$ over L in the form of an array z_{ij} for which $1 \leq i \leq J, 1 \leq j \leq Q$, where for each such i, j, $z_{ij} = Bel(\alpha_i \wedge \beta_j)$. We shall refer to

$$\{(i,j) \text{ s.t. } 1 \leq i \leq J, 1 \leq j \leq Q\} \tag{5.18}$$

as *the* $J \times Q$ *grid.*

Also, where $[z]$ represents the matrix of values of the p.b.f. Bel, then $Bel(\alpha_i) = \sum_{j=1}^{Q} z_{ij}$, which we denote by x_i for each $i = 1, \ldots J$. We say that $[z]|_{L_1} = \vec{x}$ which is $[z]$ restricted to L_1. We say that row i of $[z]$ sums to x_i. Similarly $Bel(\beta_j) = \sum_{i=1}^{J} z_{ij}$, which we denote by y_j for $j = 1, \ldots Q$. We say that column j of $[z]$ sums to y_j. $[z]|_{L_2} = \vec{y}$ is $[z]$ restricted to L_2. Note that $\vec{x} \in V^{L_1}(K_1)$ and $\vec{y} \in V^{L_2}(K_2)$.

The $Maximin$ **solutions** We let $\vec{X} = Mxmn(K_1) \in \mathbb{D}^J$ and $\vec{Y} = Mxmn(K_2) \in \mathbb{D}^Q$. Without loss of generality we assume that K_1 and K_2, with the enumerations of the atoms of L_1 and L_2 respectively, both admit the identity permutation w.r.t. Theorem 66 so that $X_1 \leq \ldots \leq X_J$ and $Y_1 \leq \ldots \leq Y_Q$.

Definition We now define recursively a specific array $[Z]$ of values Z_{ij} for $1 \leq i \leq J, 1 \leq j \leq Q$, in a sequence of *Stages*, as follows, given by Stage 1, ... Stage k_{end}. We also define sequences of integers, given by $g(1), g(2) \ldots g(k_{end})$ and $h(1), \ldots h(k_{end})$, a sequence of real numbers $N_1, \ldots N_{k_{end}}$ and a sequence $S_1, \ldots S_{k_{end}}$ of pairwise disjoint subsets of the $J \times Q$ grid.

For each $k = 1, \ldots k_{end}$ we note that for all (i, j) in the $J \times Q$ grid, we have not yet defined Z_{ij} at the start of Stage k iff $i \geq g(k)$ and $j \geq h(k)$. Then, during Stage k, we define a real number N_k and a set S_k before defining $Z_{ij} = N_k$ for all $(i, j) \in S_k$.

When the union of the S_k's defined so far is the $J \times Q$ grid, $[Z]$ is fully defined and we have completed Stage k_{end}; this process stops.

Stage 1 At the beginning of Stage 1, the set of values of (i, j) for which Z_{ij} is not yet defined is the $J \times Q$ grid so we let $g(1) = h(1) = 1$.

If $X_1/Q \leq Y_1/J$, define $Z_{1j} = X_1/Q$ for $j = 1, 2, \ldots Q$, $N_1 = X_1/Q$ and let $S_1 = \{(1, j) \,|\, 1 \leq j \leq Q\}$. Otherwise let $Z_{i1} = Y_1/J$ for $i = 1, 2, \ldots J$ and define $N_1 = Y_1/J$ and $S_1 = \{(i, 1) \,|\, 1 \leq i \leq J\}$, completing Stage 1.

After Stage k, if $\bigcup_{p=1}^{k} S_p$ is the $J \times Q$ grid, the process stops and we define

$k_{end} = k$. Otherwise we continue with:

Stage $k+1$ The set of values of (i,j) for which Z_{ij} is not yet defined is a rectangle of the form

$$\{(i,j) \text{ s.t. } G \le i \le J,\, H \le j \le Q\} \tag{5.19}$$

for some (G,H) in the $J \times Q$ grid. We define $g(k+1) = G$ and $h(k+1) = H$. Also we can see that $k+1 = g(k+1) + h(k+1) - 1$. We suppress the dependence of g, h on the Stage number for the rest of this definition. The rectangle given by (5.19) equals

$$\{(i,j) \,|\, 1 \le i \le J,\, 1 \le j \le Q\} \setminus \bigcup_{p=1}^{k} S_p \tag{5.20}$$

Let

$$Rowfin = \frac{(X_g - \sum_{j=1}^{h-1} Z_{gj})}{Q-h+1} \text{ and } Columnfin = \frac{\left(Y_h - \sum_{i=1}^{g-1} Z_{ih}\right)}{J-g+1} \tag{5.21}$$

Then if $Rowfin \le Columnfin$, define

$$S_{k+1} = \{(g,j) \text{ s.t. } h \le j \le Q\} \tag{5.22}$$

let $N_{k+1} = Rowfin$ and define $Z_{ij} = N_{k+1}$ for all $(i,j) \in S_{k+1}$. Otherwise define

$$S_{k+1} = \{(i,h) \text{ s.t. } g \le i \le J\} \tag{5.23}$$

let $N_{k+1} = Columnfin$ and define $Z_{ij} = N_{k+1}$ for all $(i,j) \in S_{k+1}$. This completes Stage $k+1$.

After Stage k_{end}, the disjoint union of all the S_k's is the $J \times Q$ grid.

Proof of Theorem 84 To do this, we show that $[Z]|_{L_1} = Mxmn(K_1)$, $[Z]|_{L_2} = Mxmn(K_2)$ and $[Z] = Mxmn(K_1 + K_2)$. Firstly we establish certain facts about $[Z]$ which can be easily illustrated by numerical examples.

Lemma 85 *If* $1 \le k < k_{end}$, $N_k \le N_{k+1}$.

Proof We fix $k = k_0$ such that $1 \le k_0 < k_{end}$. Let $G = g(k_0)$, $H = h(k_0)$ so that $k_0 = G + H - 1$. We now compare N_{k_0} and N_{k_0+1}.

Case 1(a) During Stage k_0, $Rowfin \leq Columnfin$ so $S_{k_0} = \{(G, j) \mid H \leq j \leq Q\}$. During Stage $k_0 + 1$, $Rowfin \leq Columnfin$ so $S_{k_0+1} = \{(G + 1, j) \mid H \leq j \leq Q\}$.

Then

$$N_{k_0+1} = \frac{\left(X_{G+1} - \sum_{j=1}^{H-1} Z_{(G+1)j}\right)}{Q - H + 1} \tag{5.24}$$

Consider some Z_{Gj} for which $1 \leq j \leq H - 1$. Now there exists $k' < k_0$ such that $Z_{Gj} = N_{k'}$ and the corresponding $S_{k'}$ must be a final interval of that row or that column. If it is the row then $S_{k_0+1} \subseteq S_{k'}$, so we have a contradiction as the S_k sets are disjoint. Hence, if $(G, j) \in S_{k'}$, we see that $(G + 1, j) \in S_{k'}$ so $Z_{Gj} = Z_{(G+1)j}$. Hence

$$(Q-H+1)N_{k_0+1} \;=\; X_{G+1} - \sum_{j=1}^{H-1} Z_{(G,j)} \;\geq\; X_G - \sum_{j=1}^{H-1} Z_{(G,j)} \;=\; (Q-H+1)N_{k_0} \tag{5.25}$$

since $X_{G+1} \geq X_G$ so in this case the lemma is true.

Case 1(b) During Stage k_0, $Rowfin \leq Columnfin$ so $S_{k_0} = \{(G, j) \mid H \leq j \leq Q\}$. During Stage $k_0 + 1$, $Rowfin > Columnfin$ so $S_{k_0+1} = \{(i, H) \mid G + 1 \leq i \leq J\}$.

Suppose for contradiction that $N_{k_0+1} < N_{k_0}$. Then

$$N_{k_0+1} = \frac{\left(Y_H - \sum_{i=1}^{G} Z_{iH}\right)}{J - G} \tag{5.26}$$

When the values of Z_{ij} for (i, j) in S_{k_0} were defined at Stage k_0, that was because $Z_{GH} = N_{k_0} \leq (Y_H - \sum_{i=1}^{G-1} Z_{iH})/(J - G + 1)$. Hence

$$Z_{GH}(J - G + 1) \leq Y_H - Z_{1H} - \ldots - Z_{(G-1)H} \tag{5.27}$$

but also

$$\frac{Y_H - Z_{1H} - \ldots - Z_{GH}}{J - G} < Z_{GH} \tag{5.28}$$

so it follows that $Y_H - Z_{1H} - \ldots - Z_{(G-1)H} < (J - G + 1)Z_{GH}$, which is a contradiction so $N_{k_0+1} \geq N_{k_0}$ as required.

Case 2(a) During Stage k_0, $Rowfin > Columnfin$ so $S_{k_0} = \{(i, H) \mid G \leq i \leq J\}$. During Stage $k_0 + 1$, $Rowfin \leq Columnfin$ so $S_{k_0+1} = \{(G, j) \mid H + 1 \leq j \leq Q\}$.

Case 2(b) During Stage k_0, $Rowfin > Columnfin$ so $S_{k_0} = \{(i,H)\,|\,G \leq i \leq J\}$. During Stage k_0+1, $Rowfin > Columnfin$ so $S_{k_0+1} = \{(i,H+1)\,|\,G \leq i \leq J\}$.

Cases 2(a) and 2(b) are exactly similar to Cases 1(a) and 1(b). We have proved the lemma.

■

Corollary 86 $[Z]$ *is non-negative.*

Proof For every (i,j) in the $J \times Q$ grid, there exists k such that $Z_{ij} = N_k$. By Lemma 85, $N_1 \leq N_2 \leq \ldots N_k$ and, by definition, $N_1 \geq 0$. Hence all $Z_{ij} \geq 0$ and we have proved the corollary.

■

Lemma 87 $[Z]|_{L_1} = Mxmn(K_1)$ *and* $[Z]|_{L_2} = Mxmn(K_2)$.

Proof We need to show that row i of $[Z]$ sums to X_i and column j to Y_j for each $i = 1,\ldots J, j = 1,\ldots Q$. The set S_k of those (i,j) for which Z_{ij} is defined during Stage k is either a final interval of row $g(k)$ or of column $h(k)$. The row or column completed is given the correct sum ($X_{g(k)}$ or $Y_{h(k)}$ respectively) and, apart from at the final Stage, Stage k_{end}, no other row or column is being completed.

Hence we need only check that at Stage k_{end}, every row and column being completed has the correct sum.

Let $g(k_{end}) = G$ and $h(k_{end}) = H$.

Case 1

During Stage k_{end}, $Rowfin \leq Columnfin$ so that $S_{k_{end}} = \{(J,H),\ldots(J,Q)\}$.

Then

$$N_{k_{end}} = \frac{\left(X_J - \sum_{j=1}^{H-1} Z_{Jj}\right)}{Q-H+1} \tag{5.29}$$

Now for each c s.t. $H \leq c \leq Q$ and all $i < J$, $Z_{iH} = Z_{ic}$ because (i,c) belongs to the same $S_{k'}$ as (i,H) : if it doesn't the $S_{k'}$ containing (i,c) contains (J,c) but (J,c) is in $S_{k_{end}}$; this is a contradiction since the S_k's are disjoint.

Now let

$$T = \sum_{i=1}^{J-1} Z_{iH} = \sum_{i=1}^{J-1} Z_{ic} \tag{5.30}$$

for each c s.t. $H \le c \le Q$.

Since at Stage k_{end}, $Rowfin \le Columnfin$, we see that

$$N_{k_{end}} \le Y_H - T \le Y_c - T \text{ for each } c \text{ s.t. } H \le c \le Q \tag{5.31}$$

as $\vec{Y}$ is increasing.

Hence, for each $c = H, H+1, \ldots Q$,

$$\sum_{i=1}^{J} Z_{ic} \le Y_c \tag{5.32}$$

Also since the S_k's that intersect column H must be final intervals of rows, every row of $[Z]$ has the correct sum so the sum of all Z_{ij} is $1 = \sum_{i=1}^{J} X_i$. Columns $1, \ldots H-1$ sum correctly since if $j < H$, (J, j) is in an $S_{k'}$ which is a final interval of a column; otherwise that $S_{k'}$ would intersect $S_{k_{end}}$, a contradiction since the S_k's are disjoint. Hence if there exists an integer c s.t. $H+1 \le c \le Q$ and column c sums to less than Y_c, the sum of all Z_{ij} is less than $\sum_{j=1}^{Q} Y_j$ i.e. less than 1, so we have a contradiction. Hence, by (5.32), each column of $[Z]$ must sum correctly.

Case 2

During Stage k_{end}, $Rowfin > Columnfin$ so that $S_{k_{end}} = \{(G, Q), \ldots (J, Q)\}$.

Case 2 is symmetrically similar to Case 1 and this completes the proof of the lemma.

■

Proof of Theorem 84 continued It remains to prove:

Let $[W] = Mxmn(K_1 + K_2)$. *Then for each* k *s.t.* $1 \le k \le k_{end}$, $W_{ij} = N_k$ *for each* $(i, j) \in S_k$.

We do this by induction on k.

Base Case We show that $W_{ij} = N_1$ for all $(i, j) \in S_1$.

Case 1 At Stage 1, $Rowfin \le Columnfin$ so that $S_1 = \{(1, j) \mid 1 \le j \le Q\}$ and $N_1 = X_1/Q$.

Since $[Z] \in V^L(K_1+K_2)$ (by Lemma 87) and $min([Z]) = N_1$, $min([W]) \geq N_1$ but if $W_{ij} > N_1$ for all (i,j) in the $J \times Q$ grid every row of $[W]$ sums to more than X_1, i.e. $min([W]|_{L_1}) > X_1$, even though $[W]|_{L_1} \in V^{L_1}(K_1)$. Since $\vec{X} = Mxmn^{L_1}(K_1)$, we have a contradiction. Hence $min([W]) = N_1$, and every value of $[W]$ in some row must be N_1 so that the row sums to only X_1. Now

$$min([W]|_{L_1}) = X_1 \implies ([W]|_{L_1 1}) = X_1 \tag{5.33}$$

by our assumptions on page 116 just after "The *Maximin* solutions". Hence all of the $W_{1j} = N_1$ as required.

Case 2 <u>At Stage 1, $Rowfin > Columnfin$ so that $S_1 = \{(i,1) \,|\, 1 \leq i \leq J\}$ and $N_1 = W_1/J$</u>. This case is similar to Case 1, so we have completed the Base Case.

Inductive Step Assume (I.H.) that $[W]$ agrees with $[Z]$ on $S_1, \ldots S_k$, for some k s.t. $1 \leq k < k_{end}$. $[W]$ is either equal to or *maximin*-better than $[Z]$, so by ignoring the common values at co-ordinates in $S_1, \ldots S_k$ (which we know we can do, by Corollary 75), if we let

$$R = \{(i,j) \,|\, 1 \leq i \leq J,\, 1 \leq j \leq Q\} \setminus \bigcup_{p=1}^{k} S_p \tag{5.34}$$

then $min_{(i,j)\in R} W_{ij} \geq N_{k+1}$.

Case 1 <u>At Stage $k+1$, $Rowfin \leq Columnfin$</u>.

Let $G = g(k+1)$ and $H = h(k+1)$. Also

$$N_{k+1} = \frac{\left(X_G - \sum_{j=1}^{H-1} Z_{Gj}\right)}{Q-H+1} \quad \text{and} \quad S_{k+1} = \{(G,H)\ldots(G,Q)\} \tag{5.35}$$

Let $T = \sum_{j=1}^{H-1} Z_{Gj} = \sum_{j=1}^{H-1} Z_{cj}$ for each c s.t. $c \geq G$. Suppose for contradiction that

$$min_{(i,j)\in R} W_{ij} > N_{k+1} \tag{5.36}$$

Then for each $i = G, \ldots J$,

$$[W]|_{L_1 i} > T + (Q-H+1)N_{k+1} > X_G \tag{5.37}$$

For each i less than G,

$$(i, j) \in \bigcup_{k=1}^{q} S_k \tag{5.38}$$

and $W_{ij} = Z_{ij}$ for those i and by Lemma 87, $[W]|_{L_1 i} = X_i$. Hence $[W]|_{L_1}$ is a solution of K_1 with the smallest $G - 1$ values equal to those of $\vec{X} = Mxmn(K_1)$ but the minimum of the others larger than X_G, which is a contradiction of our assumptions on page 116 just after "The *Maximin* solutions".

Thus $min_{(i,j)\in R} W_{ij} = N_{k+1}$ and some row i of $[W]$ for which $i \geq G$ must sum to exactly $X_G = T + (Q - H + 1)N_{k+1}$ or the above contradiction returns. Since

$$[W]|_{L_1} \in V^{L_1}(K_1),\ [W]|_{L_1 G} = X_G \tag{5.39}$$

so $W_{Gj} = N_{k+1}$ for each $j = H, \ldots Q$.

Case 2 At Stage $k+1$, $Rowfin > Columnfin$.

In this case $S_{k+1} = \{(G, H), \ldots (J, H)\}$ and the proof is symmetrically similar to that in Case 1, so this completes the Inductive Step and the proof of Theorem 84.

∎

5.2 Piecewise Linear Loaf Continuity of $Maximin$

Theorem 88 *Maximin is Piecewise Linear Loaf Continuous.*

Proof This can be proved just as Theorem 51, reversing some inequalities.

$\blacksquare$

5.3 Where are the discontinuities of $Maximin$?

Theorem 89 *For all $K \in CL$, $Maximin$ is continuous at K (in the Blaschke topology) $\Leftrightarrow$ $Mxmn(K)$ is the only solution of K with the maximal minimum.*

Notation In this section, we use the notation m_1 as given by (4.19).

We know that m_1 is a uniformly continuous function on CL by Lemma 76.

Proof ($\Leftarrow$) Let $K \in CL$ be fixed. We assume that $\vec{X} = Mxmn(K)$ is the only solution of K with minimum $m_1(K) = M_1$.

Claim Given fixed $\epsilon > 0$, there exists $\delta > 0$ such that for every $\vec{y} \in V^L(K)$ such that $min(\vec{y}) > m_1(K) - \delta$, $|\vec{y} - \vec{X}| < \epsilon/2$.

Proof of claim Suppose for contradiction that for some fixed positive ϵ and every $\delta > 0$ there exists $\vec{y} \in V^L(K)$ such that $\tilde{y}_1 = min(\vec{y})$ is within δ of M_1 and $|\vec{X} - \vec{y}| \geq \epsilon/2$.

For all $n \in \mathbb{N}$, let $\vec{y}^{(n)}$ be such a $\vec{y}$ for $\delta = 1/n$. By the compactness of $V^L(K)$, there exists a convergent subsequence of the $\vec{y}^{(n)}$'s whose limit, a solution of K, must have minimum M_1, by continuity of min. Hence that limit is $\vec{X}$ even though the sequence is bounded away from $\vec{X}$, which is a contradiction, so we have proved the claim.

$\blacksquare$

Now for each $\epsilon > 0$, let $\delta > 0$ be as in the above claim. Let $K' \in CL$ be such that

$$\Delta(K, K') < min\left(\frac{\delta}{2}, \frac{\epsilon}{2}\right) \tag{5.40}$$

Then, by Lemma 76, $|m_1(K') - M_1| \le \Delta(K, K') < \delta/2$ and there exists $\vec{z} \in V^L(K)$ s.t. $|\vec{z} - Mxmn(K')| < min(\frac{\delta}{2}, \frac{\epsilon}{2})$ so

$$|\,min(\vec{z}) - M_1| \;\le\; |min(\vec{z}) - m_1(K')| + |m_1(K') - M_1| \;< \delta \tag{5.41}$$

Hence by the claim, $|\vec{z} - \vec{X}| < \epsilon/2$. Also $|\vec{z} - Mxmn(K')| < \epsilon/2$ and so $|Mxmn(K') - \vec{X}| < \epsilon$. We have proved Theorem 89 in the direction ($\Leftarrow$).

($\Rightarrow$). Let $K \in CL$ be fixed such that $Mxmn(K)$ is not the only solution of K with minimum $m_1(K) = M_1$. We use the notation $m_1, m_2, \ldots m_J$ and $\tilde{}$ as in Chapter 3 and let $m_i(K) = M_i$ for each $i = 1, \ldots J$.

Case 1 $\underline{I^L(K) = \emptyset}$. The following procedure finds knowledge bases arbitrarily close to K, whose $Maximin$ values are bounded away from $Mxmn(K)$. Let

$$\vec{X} = Mxmn(K) = (M_1, M_1, \ldots M_1, M_{p+1}, \ldots M_J) \tag{5.42}$$

w.l.o.g., where $M_1 > 0$, which we can do by Theorem 78. Suppose that K admits the identity permutation w.r.t. Theorem 66. Assume also that in (5.42) $M_{p+1} > M_1$.

Lemma 90 *Let $\vec{x} \in V^L(K)$ such that $min(\vec{x}) = M_1$. Then $x_i = M_1$ for each $i = 1, \ldots p$.*

Proof Assume that $min(\vec{x}) = \tilde{x}_1 = M_1$ and $\vec{x} \in V^L(K)$. Then $x_1 = M_1$. Suppose for contradiction that $k = 1, 2$ or $\ldots$ or $p - 1$ exists such that

$$\tilde{x}_1 = x_1 = \tilde{x}_2 = \ldots = \tilde{x}_k = M_1 = M_k \tag{5.43}$$

but that $\tilde{x}_{k+1} \ne M_1$. Since $\tilde{x}_{k+1} \ge \tilde{x}_k$, $\tilde{x}_{k+1} > M_1$. Hence $x_1 = x_2 = \ldots = x_k = M_1$. The largest possible $\tilde{x}_{k+1} \ge \tilde{x}_k = M_1$ is $m_{k+1} = M_1$, so we have found a contradiction and no such k exists. Hence $\tilde{x}_i = M_1$ for each $i = 1, \ldots p$ so $x_i = M_1$ for those i and we have proved the lemma.

■

Proof of ($\Rightarrow$) **Case 1 ctd** Now, for small positive ϵ, we find knowledge bases K' such that $\Delta(K, K') < \epsilon$ but with $Mxmn(K')$ bounded away from $Mxmn(K)$ as $\epsilon \searrow 0$.

Let $\vec{Y}$ be a fixed solution of K such that $\vec{Y} \neq \vec{X}$ and $\tilde{Y}_1 = M_1$ so, by Lemma 90, $Y_1 = \ldots = Y_p = M_1$. Let the vector space

$$D = \{\lambda(\vec{z} - \vec{X}) \mid \lambda \in \mathbb{R}, \vec{z} \in V^L(K)\} \tag{5.44}$$

be generated by the orthonormal basis $\vec{u}^{(1)}, \ldots \vec{u}^{(d)}$ where $d = dim(D)$ and $\vec{u}^{(1)}$ is parallel to $\vec{Y} - \vec{X}$, so $u_1^{(1)} = u_2^{(1)} = \ldots = u_p^{(1)} = 0$. We extend this to $\vec{u}^{(1)}, \ldots \vec{u}^{(d)}, \ldots \vec{u}^{(J)}$, an orthonormal basis for $\mathbb{R}^J$.

If ϵ is small such that $\epsilon > 0$ let $\vec{U}^{(1)}$ be a unit vector in $\mathbb{R}^J - D$ chosen such that $U_1^{(1)}, U_2^{(1)}, \ldots U_p^{(1)}$ are all positive, $\sum_{i=1}^{J} U_i^{(1)} = 0$ and the angle between $\vec{u}^{(1)}$ and $\vec{U}^{(1)}$ is smaller than ϵ.

Claim There exists a vector $\vec{U}^{(1)}$ satisfying the above conditions.

Proof of claim Note that $p < J-1$, otherwise K would have only one solution $\vec{x}$ such that $x_1 = x_2 = \ldots = x_p = M_1$. Note also that if $U_1^{(1)}, U_2^{(1)}, \ldots U_p^{(1)}$ are all positive, it follows that $\vec{U}^{(1)} \notin D$ since if $\vec{U}^{(1)} \in D$, $\vec{X} + \delta\vec{U}^{(1)}$ is a solution of K for a small enough positive value of δ, but we have a contradiction since
$min(\vec{X} + \delta\vec{U}^{(1)}) > M_1$ and $\vec{X} = Mxmn(K)$.

We now show that the values of $u_i^{(1)}$ for which $p < i \leq J$ are not all equal. Suppose for contradiction that $u_{p+1}^{(1)} = u_i^{(1)}$ for all i such that $p < i \leq J$. Then, since $\sum_{i=1}^{J} u_i^{(1)} = 0$ and $u_i^{(1)} = 0$ for each $i \leq p$, we deduce that $\vec{u}^{(1)} = \vec{0}$, so we have a contradiction.

Hence w.l.o.g. we assume that $u_{p+1}^{(1)} < u_{p+2}^{(1)}$. We now show that $\vec{U}^{(1)}$ exists satisfying the above conditions such that for small enough $\delta > 0$, $U_1^{(1)} = \delta = U_i^{(1)}$ for all $i \leq p$, and $U_i^{(1)} = u_i^{(1)}$ for all i such that $p + 3 \leq i \leq J$. In other words, we can move from $\vec{u}^{(1)}$ to $\vec{U}^{(1)}$ by increasing all of the initial zeros by a small enough amount and only changing two of the other co-ordinate values.

Let $u_{p+1}^{(1)} = v$ and $u_{p+2}^{(1)} = w$, where $v < w$. Where δ is small and positive, we wish to determine values of $U_{p+1}^{(1)}$ and $U_{p+2}^{(1)}$ such that

$$U_{p+1}^{(1)} + U_{p+2}^{(1)} = v + w - p\delta \text{ and } U_{p+1}^{(1)^2} + U_{p+2}^{(1)^2} = v^2 + w^2 - p\delta^2 \tag{5.45}$$

so that $\vec{U}^{(1)}$ is a unit vector and $\sum_{i=1}^{J} U_i^{(1)} = 0$. Now, in general, if we know that σ, τ

are real such that $\sigma \leq \tau, \sigma + \tau = S$ and $\sigma^2 + \tau^2 = T$, where S and T are known real numbers, we can calculate σ and τ by

$$\sigma = \frac{S}{2} - \sqrt{\frac{T}{2} - \frac{S^2}{4}}, \quad \tau = \frac{S}{2} + \sqrt{\frac{T}{2} - \frac{S^2}{4}} \tag{5.46}$$

We can see that if $S = v + w$ and $T = v^2 + w^2$ then the expression $\frac{T}{2} - \frac{S^2}{4}$, of which we take the square root to use the above formula, is strictly positive (since $v \neq w$). Hence if δ is small enough, say less than δ_{max}, $\frac{T-p\delta^2}{2} - \frac{(S-p\delta)^2}{4}$ is also strictly positive. If we define functions V and W of δ for all $\delta \in [0, \delta_{max}]$, given by

$$\begin{aligned} V(\delta) &= \frac{v+w-p\delta}{2} - \sqrt{\frac{v^2+w^2-p\delta^2}{2} - \frac{(v+w-p\delta)^2}{4}}, \text{ and} \\ W(\delta) &= \frac{v+w-p\delta}{2} + \sqrt{\frac{v^2+w^2-p\delta^2}{2} - \frac{(v+w-p\delta)^2}{4}} \end{aligned} \tag{5.47}$$

these are continuous, so for small enough δ, $V(\delta)$ and $W(\delta)$ are as close as we need to $v = V(0)$ and $w = W(0)$ respectively. For such a value of δ, say $\delta = \delta_0$, we let

$$\vec{U}^{(1)} = (\delta_0, \delta_0, \delta_0 \ldots, V(\delta_0), W(\delta_0), u^{(1)}_{p+3}, \ldots u^{(1)}_J) \tag{5.48}$$

which is close to $\vec{u}^{(1)}$. This satisfies the required properties, so we have proved the claim.

■

Proof of ($\Rightarrow$) **Case 1 ctd** For each $j = 2, \ldots J$ define $\vec{U}^{(j)}$ to be a rotation of $\vec{u}^{(j)}$ through the same angle as for $j = 1$ such that the $\vec{U}^{(j)}$'s for $j = 1, \ldots d$ also form an orthonormal basis of a different vector space $D(\epsilon)$ and $\vec{U}^{(1)}, \ldots \vec{U}^{(J)}$ is also an orthonormal basis of $\mathbb{R}^J$. We assume w.l.o.g. that

$$K = \{(\vec{x} - \vec{X}) \cdot \vec{u}^{(j)} = 0 \text{ for } j = d+1, d+2, \ldots J\} \tag{5.49}$$

since *Maximin* satisfies Equivalence (Theorem 70). We define

$$K'(\epsilon) = \{(\vec{x} - \vec{X}) \cdot \vec{U}^{(j)} = 0 \text{ for } j = d+1, d+2, \ldots J\} \tag{5.50}$$

Lemma 91 *$\vec{X} \in V^L(K'(\epsilon))$, $D(\epsilon) = \{\lambda(\vec{z} - \vec{X}) \,|\, \lambda \in \mathbb{R}, \vec{z} \in V^L(K'(\epsilon))\}$ and as $\epsilon \searrow 0, K'(\epsilon) \to K$.*

Proof $\vec{X} \in V^L(K'(\epsilon))$ from the definition.
Also we see that $D(\epsilon) = \{\lambda(\vec{z} - \vec{X}) \mid \lambda \in \mathbb{R}, \vec{z} \in V^L(K'(\epsilon))\}$.

Sublemma 92 ([Court]) *Let $K \in CL$ and $\epsilon > 0$. Then there exists $\delta > 0$ such that if $K' \in CL$ and the matrices of coefficients of K and K' have the same number of rows and the same rank and the largest modulus difference between corresponding coefficients is less than δ and $I^L(K) = I^L(K')$ then $\Delta(K, K') < \epsilon$.*

Proof See [[Par], pp 92-94]. ■

Proof of Lemma 91 continued As $\epsilon \to 0$, $\vec{U}^{(j)} - \vec{u}^{(j)} \to 0$ for each $j > d$ so the matrix of coefficients of $K'(\epsilon)$ tends to that of K. Also $I^L(K'(\epsilon)) = I^L(K) = \emptyset$ because $\vec{X}$ is strictly positive and a solution of both $K'(\epsilon)$ and K. By definition, $rank(K) = rank(K'(\epsilon))$. Lemma 91 now follows, using Sublemma 92.

■

To show that $Mxmn(K'(\epsilon))$ is bounded away from $Mxmn(K)$ we require the following geometric result:

Sublemma 93 *Suppose that $\vec{v} \in \mathbb{D}^J$ such that, for some $dist > 0$ and some $p = 1, \dots$ or $J - 1$,*

$$dist \leq v_j - v_i \tag{5.51}$$

for all $i = 1, \dots p$, $j = p + 1, \dots J$. If $\vec{w} \in \mathbb{D}^J$ satisfies $|\vec{w} - \vec{v}| < \frac{dist}{\sqrt{2}}$ then $w_i < w_j$ for all $i = 1, \dots p$, $j = p + 1, \dots J$.

Proof Suppose for contradiction that $\vec{v}$, $\vec{w}$ satisfy the hypotheses of the sublemma but that $w_i \geq w_j$ for some $i = 1, \dots p$, $j = p + 1, \dots J$. We assume w.l.o.g. that $w_1 \geq w_J$. We know that

$$|\vec{w} - \vec{v}| < \frac{dist}{\sqrt{2}} \tag{5.52}$$

Thus, just considering the total square difference between the values at co-ordinates 1 and J,

$$|(w_1, w_J) - (v_1, v_J)| < \frac{dist}{\sqrt{2}} \tag{5.53}$$

However, in the x_1, x_J plane, straight lines for which $x_1 + x_J$ is constant are perpendicular to the line $x_1 = x_J$. By dropping a perpendicular from $v' = (v_1, v_J)$, the nearest point to v', say $w' = (w_1, w_J)$, for which $w_1 \geq w_J$ is given by

$$w' = \left(\frac{v_1 + v_J}{2}, \frac{v_1 + v_J}{2}\right) \tag{5.54}$$

and its distance from v' is

$$\frac{v_J - v_1}{\sqrt{2}} \geq \frac{dist}{\sqrt{2}} \tag{5.55}$$

so we have reached a contradiction and proved the sublemma.

■

Proof of Theorem 89 continued For small enough ϵ, $Mxmn(K'(\epsilon))$ is not near $Mxmn(K)$. Suppose for contradiction that $\vec{y} = Mxmn(K'(\epsilon))$ and

$$|\vec{y} - \vec{X}| < \frac{M_{p+1} - M_1}{\sqrt{2}} \tag{5.56}$$

Then, by Sublemma 93, $y_1, y_2, \ldots y_p$ are all less than $y_{p+1}, \ldots y_J$. Moving from $\vec{y}$ in the direction $\vec{U}^{(1)}$ would increase all of the first p co-ordinate values, increasing the minimum, so we have reached a contradiction.

However, by Lemma 91, $K'(\epsilon) \to K$ as $\epsilon \searrow 0$. Hence $Maximin$ is not continuous at K and we have proved the theorem in direction ($\Rightarrow$) in Case 1.

($\Rightarrow$) **Case 2** $\underline{I^L(K) \neq \emptyset}$. We assume w.l.o.g. that $I^L(K) = \{1, 2, \ldots p\}$ so that $Mxmn(K) = (0, 0, \ldots 0, M_{p+1} \ldots M_J)$ where $M_{p+1} > 0$, and we can do this by Theorem 78. Thus $x_i = 0$ for all $\vec{x} \in V^L(K), i = 1, \ldots p$. We also write, w.l.o.g.,

$$V^L(K) = \{\vec{x} = (0, 0, \ldots 0, \lambda_1, \lambda_2 \ldots \lambda_d, F_1, F_2, \ldots F_{J-p-d}) \mid \vec{\lambda} \in \Lambda, \} \tag{5.57}$$

where for each $h = 1, \ldots J - p - d$,

$$F_h = \sum_{i=1}^{d} F_{hi}\lambda_i + c_h \tag{5.58}$$

is a linear function of the parameters $\lambda_1, \ldots \lambda_d$, $d \geq 1$ and Λ is the set of values of $(\lambda_1, \ldots \lambda_d)$ such that $\vec{x}$, as in (5.57), is non-negative. We assume w.l.o.g. that λ_1 is not constant in Λ.

Let ϵ be small and positive; then we may, by Lemma 4, define K_ϵ such that

$$V^L(K_\epsilon) = \{\vec{x} = (\epsilon\lambda_1, 0, \ldots 0, \lambda_1(1-\epsilon), \lambda_2, \ldots \lambda_d, F_1, F_2 \ldots F_{J-p-d}) \,|\, \vec{\lambda} \in \Lambda\} \quad (5.59)$$

We can see that in the above expression for $\vec{x}$, $\vec{x}$ is non-negative iff $(\lambda_1, \ldots \lambda_d) \in \Lambda$. Now $\Delta(K, K_\epsilon) \leq 2\epsilon$ since the distance between an $\vec{x}$ as in (5.59) and $\vec{x}$ as in (5.57) using *the same* $(\lambda_1, \ldots \lambda_d)$ for each, is not more than 2ϵ.

Suppose for contradiction that

$$Mxmn(K_\epsilon) = \vec{y} \text{ and } |\vec{y} - \vec{X}| < \frac{M_{p+1}}{\sqrt{2}} \quad (5.60)$$

Then, by Sublemma 93, y_1 is less than each of $y_{p+1}, \ldots y_J$ but if we replace the $(\lambda_1, \ldots \lambda_d) \in \Lambda$ that gives $\vec{y}$ in (5.59) with $(\lambda'_1, \ldots \lambda'_d) \in \Lambda$ such that $\lambda'_1 > \lambda_1$, we have a solution of K_ϵ with a larger minimum than $min(\vec{y})$ - this is a contradiction. Hence $Mxmn(K_\epsilon)$ is bounded away from $Mxmn(K)$ and we have proved the theorem.

■

Chapter 6

Algorithms for calculating *Minimax* and *Maximin*

In this chapter K will usually be a consistent constraint set on J co-ordinates. Recall from the definition in the Introduction that, although $\sum_{i=1}^{J} x_i = 1$ need not be a constraint of K, the set of non-negative solutions $V(K)$ must be bounded. $Minimax(K)$ and $Maximin(K)$ are well-defined with almost exactly the same proofs as before: Lemma 27, Theorem 28 and Corollary 29 for $Minimax$ and Lemma 65, Theorem 66 and Corollary 67 for $Maximin$ go through for constraint sets. We also use Theorem 37, Theorem 78 and Theorem 72 with constraint sets in general.

We extend the Blaschke metric and definitions of a loaf and Piecewise Linear Loaf Continuity to constraint sets in the obvious way. Lemma 76 generalises similarly, so Theorem 88 carries over to this general case.

Notation If J, d are integers such that $0 \leq d \leq J$ and $J \geq 2$, we let $\mathcal{C}_J^d$ denote the set of consistent constraint sets K on J co-ordinates such that $rank(K) = J - d$ (so $dim(G(K)) = d$). We use this definition for the rest of this thesis.

Notation An *E-constraint* is a constraint of the form $(x_i = x_j)$ for some i, j and we denote it by E_{ij}. Recall from Chapter 3 that a Z-constraint is a constraint of the form $(x_i = 0)$. A constraint is an *EZ-constraint* if it is either an E-constraint or a Z-constraint. This applies for the rest of this thesis.

Recall from the Introduction the definition of the generalised solutions of K, $G(K)$ where K is a constraint set. Note that $G(K)$ is not a function of $V(K)$ since the constraint sets $\{x_1 = 0, x_2 = 0\}$ and $\{x_1 + x_2 = 0\}$ have the same solutions but not the same generalised solutions.

In this chapter we shall consider affine subsets of $\mathbb{R}^J$, i.e. sets of the form:

$$\left\{ \vec{a} + \sum_{\tau=1}^{d} \tau_i \vec{dir}^{(i)} \text{ s.t. } \tau_i \in \mathbb{R} \text{ for } i = 1 \dots d \right\} \tag{6.1}$$

for some $\vec{a}, \vec{dir}^{(1)}, \vec{dir}^{(2)} \dots \vec{dir}^{(d)}$ and we denote (6.1) by $(\vec{a} + < \vec{dir}^{(1)}, \dots \vec{dir}^{(d)} >)$.

If $K \in \mathcal{C}_J^d$, $G(K)$ is an affine hyperplane of dimension d, which can be characterised by

$$G(K) = (\vec{a} + < \vec{u}^{(1)}, \dots \vec{u}^{(d)} >) \tag{6.2}$$

where the $\vec{u}^{(n)}$ are fixed linearly independent real vectors in $\mathbb{R}^J$.

For $K \in \mathcal{C}_J^d$, let $D(K)$ be the linear subspace of $\mathbb{D}^J$ given by

$$D(K) = \{\vec{y} - \vec{x} | \, \vec{x}, \, \vec{y} \in G(K)\} \tag{6.3}$$

Hence, if $G(K)$ takes the form (6.2), the vectors $\vec{u}^{(n)}$ for $n = 1, \dots d$ together form a linear basis for $D(K)$ and we say that members of $D(K)$ are *parallel* to $G(K)$.

6.1 Calculating $Minimax$

In this section of this chapter, where we write $m_1, m_2, \dots m_J$ and $\tilde{*}$, we are using the notation in the sense of Chapter 2.

Algorithm 94 (*$Minimax$* Calculation Algorithm)

Input: *Some* $K \in \mathcal{C}_J^d$.

Output: $Mmx(K)$.

- (1) For every constraint set of the form

$$K + \; d \;\; EZ\text{-constraints} \tag{6.4}$$

which has rank J, calculate its unique generalised solution and collect these solutions into a set called $Candidates$.

- (2) Delete from $Candidates$ all vectors which have negative values.

- (3) Now find $min\{(max(\vec{x}))$ s.t. $\vec{x} \in Candidates\}$. Cast out anything from $Candidates$ with greater maximum than this.

- (4) Repeat Step 3, but now retaining only those members of $Candidates$ with the smallest i'th largest value, where at each iteration i increases by 1 through the values 2,3, ... until just one vector remains in $Candidates$.

- (5) Output that vector and stop.

Proof We first prove a result in a similar spirit to Lemma 42 which can help in calculating $Minimax$.

In general, a constraint set K might have some constant co-ordinates, and/or force certain distinct co-ordinates to take the same value. Here it is assumed that K has more than one solution.

When calculating $Minimax$ these constants and equivalences can be ignored in the following precise sense:

Let the K-constant co-ordinates be $c_1 \dots c_k$. For each $i = 1, \dots k$, we define C_i such that $x_{c_i} = C_i$ for all $\vec{x} \in V(K)$ and let $C = \sum_{i=1}^{k} C_i$.

Definition The equivalence relation K-*equivalence* on $\{1, 2, \dots J\} \setminus \{K\text{-constant co-ordinates}\}$, denoted by $\sim_K$, is given by

$$i \sim_K j \iff x_i = x_j \text{ for all } \vec{x} \in V(K) \tag{6.5}$$

where the equivalence classes w.r.t. $\sim_K$ are given by $[e_1], \dots [e_Q]$ and $Q \geq 2$ since K has more than one solution.

We now define the <u>equality-constant simplification of K</u>, written $ECS(K)$, as follows:

In the constraints of K, for every $c_1 \dots c_k$ we substitute C_i for appearances of x_{c_i}, and substitute y_p for appearances of x_q when $q \sim_K e_p$. The resulting set of equations is $ECS(K)$. The notation y is chosen just so as to be distinct from x.

Now we define

$$ecs_K : V(K) \to V(ECS(K)) \text{ by } ecs_K(\vec{x}) = (x_{e_1}, \dots x_{e_Q}) \tag{6.6}$$

Lemma 95 *$ecs_K : V(K) \to V(ECS(K))$ is well-defined and is a bijection whose inverse we denote by ecs_K^{-1}.*

Proof If $\vec{x}$ satisfies K, by checking the constraints individually we see that $ecs_K(\vec{x})$ satisfies $ECS(K)$. Non-negativity of $ecs_K(\vec{x})$ is also clear. Given $\vec{y} \in V(ECS(K))$, we can find a unique $\vec{X}$ such that $ecs_K(\vec{X}) = \vec{y}$. Consider each co-ordinate $1, \dots J$ in turn. If i is K-constant then $X_{c_i} = C_i$ is forced. Otherwise there exists a unique $p = 1, \dots Q$ such that $i \sim_K e_p$. This forces $X_i = y_p$. There is no choice at any step in this procedure, which does find such an $\vec{X}$. Hence ecs_K is a bijection and we know how to calculate its inverse. We have proved the lemma.

■

We see that the operation ECS respects equivalence, i.e.

$$K \equiv K' \text{ implies } ECS(K) \equiv ECS(K') \tag{6.7}$$

where it is understood that the equivalence classes of variables of K' may have different representatives from those of K, but that this just gives us different labels of the variables y_i. We use constraint sets in this chapter since even if K is a knowledge base, K' has Q variables which might not sum to one at every solution, and Q is possibly not a power of 2.

Lemma 96 $ecs_K(Mmx(K)) = Mmx(ECS(K))$

Proof Let $\vec{X} = Mmx(K) \in \mathbb{R}^J$ and let $\vec{Y} = Mmx(ECS(K)) \in \mathbb{R}^Q$, where $\vec{a} = ecs_K(\vec{X}) \in \mathbb{R}^Q$ and $\vec{b} = ecs_K^{-1}(\vec{Y}) \in \mathbb{R}^J$. We assume w.l.o.g. that $ECS(K)$ admits the identity permutation w.r.t. Theorem 28.

We show by induction on p for $p \leq Q$ that

Claim(p) $\tilde{a}_1 = \tilde{Y}_1 \dots \tilde{a}_p = \tilde{Y}_p$ so that $Y_1 = a_1 \dots Y_p = a_p$.

Base Case $\underline{p = 0}$ In this case the claim is trivial.

Inductive Step Assuming (I.H.) that the largest p values in $\vec{a}$ match the largest p values in $\vec{Y}$, then $\tilde{X}_k = \tilde{b}_k$ for each k not greater than

$$FIX = \left(\sum_{i=1}^{p} |[e_i]| \right) + |\{j \,|\, C_j \geq a_p\}| \tag{6.8}$$

Suppose for contradiction that $\tilde{Y}_{p+1} \neq \tilde{a}_{p+1}$. Then $\tilde{Y}_{p+1} < \tilde{a}_{p+1}$ by definition of $Minimax$ and $\tilde{X}_{FIX+1}$ is larger than $\tilde{b}_{FIX+1}$, but this contradicts the definition of $\vec{X}$. Hence $\tilde{Y}_{p+1} = \tilde{a}_{p+1}$ so $Y_{p+1} = a_{p+1}$ and the Inductive Step is proved, completing the claim and the proof of the lemma.

■

Remarks Note that although we can define $Ren_r(ECS(K))$ and $ME(ECS(K))$ in similar ways, the statements corresponding to Lemma 96 are not true.

Actually the function ecs_K does NOT preserve the $minimax$ ordering. However the fact that the co-ordinates where the top p values occur match (due to the fixing of the $\sigma(1), \sigma(2)..$) between $\vec{Y}$ and $\vec{a}$ plays a crucial role, since then the "bunches of equivalent atoms" (if K is a knowledge base) in the original setting will be of matching size.

Proof of Algorithm 94 continued The following lemma suggests that it is indeed useful for us to add EZ-constraints to a constraint set K in order to calculate $Mmx(K)$.

Lemma 97 *Let K be a consistent constraint set which does not have any constant co-ordinates and whose relation $\sim_K$ is equality. Either*

$m_J(K) = 0$ *or*

$m_1(K) = m_2(K)$.

Proof We can assume, w.l.o.g., that K admits the identity permutation w.r.t. Theorem 28, so that $Mmx(K) = (m_1, m_2 \ldots m_J)$. We will suppress the dependence of the m_i's on K in this proof.

Suppose for contradiction that neither $m_J = 0$ nor $m_1 = m_2$. Then

$$m_1 > m_2 \geq \ldots \geq m_J > 0 \tag{6.9}$$

Since the co-ordinate 1 is not K-constant, there exists $\vec{u}$ in $D(K)$ such that $u_1 \neq 0$. Otherwise the 1st co-ordinate would have value m_1 everywhere in $G(K)$, contradicting our assumptions. Now we define P by

$$P = \{\vec{x} \in \mathbb{R}^J \text{ s.t. } x_1 > x_j \text{ for all } j = 2, \dots J\} \tag{6.10}$$

Since

$$(Mmx(K) + <\vec{u}>) \subseteq G(K) \tag{6.11}$$

and $Mmx(K)$ is *strictly* positive, then

$$\{Mmx(K) + \delta\vec{u} \mid \delta \in (-\epsilon, \epsilon)\} \subseteq V(K) \cap P \tag{6.12}$$

for sufficiently small $\epsilon > 0$. This is because the set of strictly positive vectors is open, as is P. Without loss of generality $u_1 < 0$. Then $Mmx(K) + \frac{\epsilon}{2}\vec{u}$ is a solution of K with the 1st co-ordinate value maximal but smaller than m_1, which is a contradiction. We have proved the lemma.

■

Lemma 98 *For any constraint set K with more than one solution, there does not exist a $ECS(K)$-constant co-ordinate, nor do there exist distinct i, j such that $y_i = y_j$ for all $\vec{y} \in V(ECS(K))$.*

If K is a set of constraints on $x_1, \dots x_J$ and $ECS(K)$ a set of constraints on $y_1, \dots y_Q$ then if a co-ordinate is $ECS(K)$-constant, say 1 w.l.o.g., then every co-ordinate $i \in [e_1]$ is K-constant but only the non-constant co-ordinates of K are affected by $\sim_K$, so we have a contradiction. Similarly if, w.l.o.g., $y_1 = y_2$ for every $\vec{y} \in V(ECS(K))$ then the two corresponding equivalence classes of co-ordinates of K $[e_1], [e_2]$, must be s.t. $x_{e_1} = x_{e_2}$ for every $\vec{x} \in V(K)$, so actually $e_1 \sim_K e_2$, again giving a contradiction so we have proved the lemma.

■

The following theorem takes us closer to proving Algorithm 94.

Theorem 99 *Let K be a constraint set with more than one solution. Then either $Mmx(K)_i = 0$ for all co-ordinates i in some equivalence class of non-constant co-ordinates, or there exist co-ordinates a, b such that $Mmx(K)_a = Mmx(K)_b$ but not $a \sim_K b$.*

Proof $Mmx(ECS(K)) = ecs_K(Mmx(K))$ by Lemma 96. $ECS(K)$ has no constant co-ordinates or distinct co-ordinates forced to be equal by Lemma 98 so, by Lemma 97, either $m_1(ECS(K)) = m_2(ECS(K))$ or $m_J(ECS(K)) = 0$. Using the fact that $Mmx(K) = ecs_K^{-1} Mmx(ECS(K))$ completes the proof of the theorem.

■

Proof of Algorithm 94 continued Firstly observe that as long as $Mmx(K) \in Candidates$ at Step 1, then it will be the output. This is because once the partly negative vectors are ignored, $Candidates$ is a subset of $V(K)$ and $Mmx(K)$ is the *minimax*-best vector in $V(K)$. Hence it does not matter what else is in $Candidates$.

Lemma 100 *For each $q = 0, 1, \ldots d$, there exist EZ-constraints $con_1, \ldots con_q$ such that $Mmx(K)$ is a solution of*

$$K + \{con_1, con_2, \ldots con_q\} \tag{6.13}$$

and the rank of (6.13) *is $J - d + q$.*

Proof We do this by induction on $q \leq d$.

Base Case $\underline{q = 0}$ In this case the lemma is trivial.

Inductive Step Suppose (I.H.) that the lemma is true in the case $q = m$ such that $m < d$. For the case $q = m + 1$ one extra EZ-constraint will be added. Given that

$$K_m = K + \{con_i \,|\, 1 \leq i \leq m\} \tag{6.14}$$

is the constraint set corresponding to (6.13) when $q = m$, there are two cases:

Case 1 $\underline{V(K_m) \text{ has more than one member}}$.

By the Inductive Hypothesis and Obstinacy (Theorem 37 for constraint sets) $Mmx(K) = Mmx(K_m)$. Not every co-ordinate is K_m-constant so, by Theorem 99, there exists an EZ-constraint, say con_{m+1}, which is satisfied by $Mmx(K)$ but which is not satisfied by every solution of K_m. Let $K_{m+1} = K_m + con_{m+1}$. Then $G(K_{m+1})$ is a strict subset of $G(K_m)$. We increase the rank of the constraint set by 1 when we add con_{m+1} to K_m and the Inductive Step is proven in this case.

Case 2 $\underline{V(K_m) = \{Mmx(K)\}}$.

Since $m < d$ the rank of the constraint set K_m is less than J so $G(K_m)$ is an affine hyperplane with strictly positive dimension. Then:

Claim There exists a co-ordinate with value zero at $Mmx(K)$ which is not zero throughout $G(K_m)$.

Proof of claim Suppose for contradiction that the claim does not hold. We fix a direction vector $\vec{u} \in D(K_m)$, i.e. parallel to $G(K_m)$, such that for all i, if $Mmx(K)_i = 0$, $u_i = 0$. Since the other values at $Mmx(K)$ are strictly positive,

$$Mmx(K) + \epsilon\vec{u} \tag{6.15}$$

is non-negative if $\epsilon \in \mathbb{R}$ is small enough. This is similar to the proof of Lemma 97. Hence there exist non-negative solutions of K_m apart from $Mmx(K)$, so we have proved a contradiction and the claim follows.

■

Hence we can choose a Z-constraint, say con_{m+1} such that when we let

$$K_{m+1} = K_m + con_{m+1} \tag{6.16}$$

then $V(K_{m+1}) = \{Mmx(K)\}$ and $rank(K_{m+1}) = rank(K_m) + 1$. We have completed the proof of this case and of the Inductive Step, so the lemma is proved.

■

Finally, to complete the proof of Algorithm 94

By Lemma 100, setting $q = d$, $Mmx(K) \in Candidates$ at Step 1, so the sorting process that follows in Steps 2-5 leads to the algorithm terminating and outputting $Mmx(K)$. We have proved the algorithm.

■

6.2 Calculating $Maximin$

In this section, when we write $m_1, \ldots m_J$ and $\tilde{*}$ we are using the notation in the sense of Chapter 4. We find algorithms that calculate $Maximin$ just as fast, indeed a little faster, than Algorithm 94 calculates $Minimax$. We use the notion of $ECS(K)$, the equality-constant simplification of K, defined in Section 6.1.

Algorithm 101 ($Maximin$ Calculation Algorithm Prototype)

***Input**: Some $K \in \mathcal{C}_J^d$.*

***Output**: $Mxmn(K)$.*

- (1) For every constraint set of the form

$$K + \ d \ EZ\text{-constraints} \tag{6.17}$$

  which has rank J, calculate its unique generalised solution and collect these solutions into a set called $Candidates$.

- (2) Cast out from $Candidates$ all vectors which have negative values.

- (3) Now find $max\{min(\vec{x}) \text{ s.t. } \vec{x} \in Candidates\}$. Cast out anything from $Candidates$ with smaller minimum than this.

- (4) Repeat Step 3, but now retaining only those members of $Candidates$ with the largest i'th smallest value, where at each iteration i increases by 1 through the values 2,3, ... until just one vector remains in $Candidates$.

- (5) Output that vector and stop.

Proof This proof is similar to that of Algorithm 94.

Lemma 102 $ecs_K(Mxmn(K)) = Mxmn(ECS(K))$.

Proof We can prove this exactly as we proved Lemma 96 but with some inequalities reversed.

$\blacksquare$

Lemma 103 *If K is a constraint set which does not have any constant co-ordinates and whose relation $\sim_K$ is equality, then $m_1 = m_2$.*

Proof If not then $m_1 < m_2$. Suppose w.l.o.g. that K admits the identity permutation w.r.t. Theorem 66. Now $I(K) = \emptyset$ so we can write

$$Mxmn(K) = (m_1, m_2, \ldots m_J) \tag{6.18}$$

where $0 < m_1 \leq \ldots \leq m_J$, by Open-mindedness (Theorem 78 for constraint sets).

Since the co-ordinate 1 is not K-constant there exists a direction parallel to $G(K)$, $\vec{u}$, s.t. $u_1 \neq 0$, say $u_1 > 0$ w.l.o.g.. Then for small enough $\epsilon > 0, \vec{y} = Mxmn(K) + \epsilon\vec{u}$ is strictly positive (as $Mxmn(K)$ is) and its minimum y_1 is greater than m_1. This is a contradiction since $\vec{y} \in V(K)$, so the lemma is proved.

■

Theorem 104 *Let K be a constraint set with more than one solution. Then there exist co-ordinates a, b such that $Mxmn(K)_a = Mxmn(K)_b$ but not $a \sim_K b$.*

Proof We state below the result for $Maximin$ which is the analogue of Theorem 99:

Let K be a constraint set with more than one solution. Then either $Mxmn(K)_i = 0$ for all co-ordinates i in some equivalence class of non-constant co-ordinates, or there exist co-ordinates a, b such that $Mxmn(K)_a = Mxmn(K)_b$ but not $a \sim_K b$.

We can prove this by following the methods of Section 6.1, reversing some inequalities. However, Theorem 104 then follows, since $Maximin$ satisfies Open-mindedness (Theorem 78 for constraint sets) so $Mxmn(K)_i > 0$ if i is not a K-constant co-ordinate.

■

Proof of Algorithm 101 The rest of the proof is similar to that of Algorithm 94 with some inequalities reversed.

■

Remark The $Maximin$ Calculation Algorithm Prototype is merely the result of reversing some inequalities in the steps of the $Minimax$ Calculation Algorithm. As the name suggests we can somewhat refine this. We do this in the following section.

6.3 Improving the Prototype Algorithm

Notation In this section and the rest of the chapter, we look at algorithms which calculate $Maximin$ and prove some theorems which tell us ways of deciding whether $\vec{X} = Mxmn(K)$. It is useful for us to characterise $Maximin$ as follows:

$$\text{For all } K \in \mathcal{C}_J^d,\ Mxmn(K) \text{ is the unique } \vec{x} \in G(K) \text{ which is } maximin\text{-best} \tag{6.19}$$

since the generalised solutions of K which are not solutions of K have negative minima anyway.

Where an affine set $S = (\vec{a}+ < \vec{dir}^{(1)}, \ldots \vec{dir}^{(d)} >)$ is such that there exists a unique point in S, say $\vec{x}$, which is *maximin*-best, we can write

$$\vec{x} = Mxmn(\vec{a}+ <\vec{dir}^{(1)}, \ldots \vec{dir}^{(d)}>) \tag{6.20}$$

In particular, if $S = G(K)$ for some $K \in \mathcal{C}_J^d$, then $\vec{x} = Mxmn(K) = Mxmn(S)$.

From now on in this chapter, we often use $G(K)$ instead of $V(K)$ to take advantage of (6.19). In light of this, we make the following definitions.

Definition For any $K_1, K_2 \in \mathcal{C}_J^d$, we say that K_1 is *gen-equivalent* to K_2 iff $G(K_1) = G(K_2)$. In this case, we write

$$K_1 \equiv_G K_2 \tag{6.21}$$

Recall that $K_1 \equiv K_2$ iff $V(K_1) = V(K_2)$. We say that K_1 *gen-implies* K_2 iff $G(K_1) \subseteq G(K_2)$ (iff $K_1 + K_2 \equiv_G K_1$). This is denoted by

$$K_1 \Rightarrow K_2 \tag{6.22}$$

Otherwise $K_1 \not\Rightarrow K_2$. In particular, if for some E-constraint E_{ij}, $K_1 \not\Rightarrow E_{ij}$ this implies that, if $K_1 + E_{ij}$ is consistent, $rank(K_1 + E_{ij}) = rank(K_1) + 1$ and E_{ij} is linearly independent of K_1.

Notation In practice, when we use this notation, there is usually a specific vector which satisfies both K, K_1 etc. and the E-constraints added to it, so we do not state consistency explicitly.

Back to improving on the $Maximin$ **Calculation Algorithm Prototype**

We first note that, in Algorithm 101, Step 2 isn't necessary since we can characterise $Maximin$ by (6.19); if a vector has negative values it is cast out at Step 3 anyway. Suppose that $K \in \mathcal{C}_J^d$. For each $i = 1, \ldots J$,

$$Mxmn(K)_i = 0 \Longrightarrow i \in I(K) \tag{6.23}$$

by Theorem 78 (for constraint sets), so it seems unnecessary for the algorithm to add Z-constraints to K at Step 1. When we inspect the proof of Algorithm 101, we find that Z-constraints come in only because after m E-constraints are added it is possible that $V(K_m) = \{Mxmn(K)\}$ but the dimension of $G(K_m)$ is still at least 1.

However, even in that scenario, we can still show that there exists an E-constraint which is satisfied by $Mxmn(K) = Mxmn(K_m)$ but is linearly independent of K_m. Thus we can simplify Step 1.

Notation If $\vec{a}, \vec{dir} \in \mathbb{R}^J$, $lev \in \mathbb{R}$, $\vec{dir}$ has *sign conflict at lev* w.r.t. $\vec{a}$ means that there exist distinct i, j s.t. $a_i = a_j = lev$, dir_i, dir_j are non-zero and of opposite sign to each other and for each k such that $a_k < lev$, $dir_k = 0$.

In other words, as we move from $\vec{a}$ in the direction $\vec{dir}$, of the co-ordinates of least value that are moving, some go up and others down.

Lemma 105 *Let* $\vec{a}$, $\vec{dir} \in \mathbb{R}^J$ *such that* $\vec{dir} \neq \vec{0}$. *Let* $lev \in \mathbb{R}$ *be the minimal value* a_i *such that* $dir_i \neq 0$. *Then the following are equivalent:*

(i) $Mxmn(\vec{a} + <\vec{dir}>)$ *exists and equals* $\vec{a}$.

(ii) $\vec{dir}$ *has sign conflict at lev w.r.t.* $\vec{a}$.

Proof ($\Leftarrow$) Suppose that $\vec{dir}$ has sign conflict at lev w.r.t. $\vec{a}$, then we can assume w.l.o.g. that

$$\vec{a} = (a_1, a_2, \ldots a_k = lev = a_{k+1} = \ldots a_p, a_{p+1} \ldots a_J) \tag{6.24}$$

where $a_1 \leq a_2 \ldots \leq a_J$, $dir_k < 0, dir_{k+1} > 0$ and $a_i = lev \Leftrightarrow k \leq i \leq p$. Also $dir_i = 0$ for each i s.t. $i < k$.

To obtain $Mxmn(\vec{a} + < \vec{dir} >)$ we can use Corollary 75 to see that we can ignore the co-ordinates i, s.t. i is less than k, since they are constant in value along the line. Every $\vec{a} + \tau\vec{b}$ for which $\tau \neq 0$ has smaller minimum than lev, that of $\vec{a}$, as either co-ordinate k or $k+1$ has a smaller value (depending on sign of τ).

($\Rightarrow$): Suppose that $\vec{a} = Mxmn(\vec{a} + < \vec{dir} >)$. Then if every dir_i such that $a_i = lev$ has the same sign, a small non-zero τ exists such that for some integer $q < J$, $\vec{a} + \tau\vec{dir}$ has the same q smallest values as $\vec{a}$ but all other values greater than lev so $\vec{a} \neq Mxmn(\vec{a} + < \vec{dir} >)$, which is a contradiction. Hence we have proved the lemma.

■

Notation Suppose that $\vec{a}, \vec{dir} \in \mathbb{R}^J$ are such that $\vec{a} \neq Mxmn(\vec{a} + <\vec{dir}>)$. Let $lev \in \mathbb{R}$ be the minimal value a_i such that $dir_i \neq 0$. By Lemma 105, all dir_i such that $a_i = lev$ have the same sign or equal zero. If they are all non-negative,

$$\vec{a} + \epsilon\vec{dir} \tag{6.25}$$

is *maximin*-better than $\vec{a}$ if ϵ is small enough and positive. We say that $\vec{dir}$ *produces maximin-improvement from* $\vec{a}$. Otherwise the dir_i such that $a_i = lev$ are all non-positive, and $-\vec{dir}$ produces *maximin*-improvement from $\vec{a}$. We use this notation in the rest of this thesis.

Lemma 106 *Suppose that $\vec{b} = Mxmn(Li)$, where $\vec{b} \in \mathbb{R}^J$ and $Li = (\vec{b} + <\vec{dir}>)$ for some $\vec{dir} \in \mathbb{R}^J$. Then if $\vec{a}$, $\vec{c} \in Li$ are such that $\vec{a}$ is a convex combination of $\vec{b}$ and $\vec{c}$, $\vec{a}$ is maximin-better than $\vec{c}$.*

Remark This lemma says that as we approach the *maximin*-best point of a line from a particular direction, our location becomes *maximin*-better.

Proof We may assume w.l.o.g., by Lemma 105, that $\vec{dir}$ produces *maximin*-improvement from $\vec{a}$, and that

$$\vec{a} = (a_1, a_2, \ldots a_k = lev = a_{k+1} = \ldots a_p, a_{p+1} \ldots a_J) \tag{6.26}$$

where $a_1 \leq a_2 \ldots \leq a_J$, $a_i = lev \Leftrightarrow k \leq i \leq p$, $dir_k > 0$ and $dir_i \geq 0$ for each i s.t. $k \leq i \leq p$. For all $\tau \in \mathbb{R}$, we define

$$\vec{x}(\tau) = \vec{a} + \tau\vec{dir} \tag{6.27}$$

which describes every point on Li exactly once.

If $\tau < 0$, the values of some of the $x_i(\tau)$ such that $a_i = lev$ are less than lev. Since co-ordinates $1, \ldots k-1$ are constant in value on Li, by Corollary 75, we see that $\vec{a}$ is *maximin*-better than $\vec{x}(\tau)$ for all negative τ.

Now $\vec{b}$ is *maximin*-better than $\vec{a}$, by definition. Hence, when we write $\vec{b} = \vec{x}(\tau_b)$, $\tau_b > 0$, otherwise we would have a contradiction. However, by our assumptions, $\vec{c}$ is on the other side of $\vec{a}$ so $\vec{c} = \vec{x}(\tau_c)$, where $\tau_c < 0$ and we deduce that $\vec{a}$ is *maximin*-better than $\vec{c}$, as required. We have proved the lemma.

■

Theorem 107 *Run Maximin Calculation Algorithm Prototype with input $K \in \mathcal{C}_J^d$. Then if only E-constraints are allowed at Step 1, the output is still $Mxmn(K)$.*

Proof By the discussion on page 141, we assume that $K \in \mathcal{C}_J^d$ and K_m is of the form

$$K_m = K + \{con_i \,|\, 1 \leq i \leq m\} \tag{6.28}$$

where the con_i are E-constraints, $Mxmn(K_m) = Mxmn(K) = \vec{X}$ and $V(K_m) = \{\vec{X}\}$. We need only show that there exists an E-constraint, say E_{ij}, satisfied by $\vec{X}$ such that $K_m \not\Rightarrow E_{ij}$. Let $\vec{dir} \in D(K_m)$. Then

$$\vec{X} = Mxmn(\vec{X} + <\vec{dir}>) \tag{6.29}$$

and by Lemma 105 $\vec{dir}$ has sign conflict at lev w.r.t. $\vec{X}$, for some $lev \in \mathbb{R}$. Hence we can fix co-ordinates $i*, j*$ such that $X_{i*} = X_{j*} = lev$ and $dir_{i*} > 0, dir_{j*} < 0$. Hence

$\vec{X}$ is the only point on the line $(\vec{X} + <\vec{dir}>)$ that satisfies the E-constraint:

$$x_{i*} = x_{j*} \tag{6.30}$$

which completes the proof of the theorem.

■

Algorithm 108 (*Maximin* Calculation Algorithm 1)

Input: *Some* $K \in \mathcal{C}_J^d$.

Output: $Mxmn(K)$.

- (1) For every constraint set of the form

$$K + \; d \; E\text{-constraints} \tag{6.31}$$

  which has rank J, calculate its unique generalised solution and collect these solutions into a set called $Candidates$.

- (2) Now find $max\{min(\vec{x})$ s.t. $\vec{x} \in Candidates\}$. Cast out anything from $Candidates$ with smaller minimum than this.

- (3) Repeat Step 2, but now retaining only those members of $Candidates$ with the largest i'th smallest value, where at each iteration i increases by 1 through the values 2,3, ... until just one vector remains in $Candidates$.

- (4) Output that vector and stop.

Proof of algorithm This follows from the remarks after the proof of Algorithm 101 and from Theorem 107.

■

6.4 Calculating *Maximin* faster and using Principal Sightings

Remarks Algorithm 108 could be regarded as rather primitive since the results of all possible ways of adding E-constraints to the constraint set K are compared in the *maximin* ordering. Surely we can calculate *Maximin* by testing fewer sets of extra constraints!

One improvement is suggested by the observation that, having looked at

$$K + \{x_1 = x_2, x_1 = x_3\} \tag{6.32}$$

we do not need to inspect

$$K + \{x_1 = x_2, x_2 = x_3\} \text{ or } K + \{x_1 = x_2, x_2 = x_3, x_1 = x_3\} \tag{6.33}$$

This suggests that we look at the possible equivalence relations being added to K, not at lists of separate constraints. If we run Algorithm 108 with input $K \in \mathcal{C}^d_J$, we add d E-constraints to K, say $\mathcal{E} = \{E_{i_1 j_1} \dots E_{i_d j_d}\}$ to get a system of rank J. Thus $rank(\mathcal{E}) = d$.

Notation For a set of E-constraints $\mathcal{E}$, we define $\sim_{\mathcal{E}}$ to be the minimal equivalence relation on $\{1, 2, \dots J\}$ for which $i \sim j$ for each $E_{ij} \in \mathcal{E}$.

Theorem 109 *For each $d = 0, 1, \dots J$ and all sets $\mathcal{E}$ of E-constraints on $\{x_1, \dots x_J\}$, $rank(\mathcal{E}) = d \iff \sim_{\mathcal{E}}$ has $J - d$ equivalence classes.*

Proof For any such $\mathcal{E}$, the set S of vectors $\vec{x} \in \mathbb{R}^J$ which satisfy $\sim_{\mathcal{E}}$ is the same as the set of vectors that satisfy $\mathcal{E}$ which is an affine set. The number of dimensions of S is the number of equivalence classes of $\sim_{\mathcal{E}}$, which must equal $J - rank(\mathcal{E})$. We have proved the theorem.

■

Notation It is convenient for us to abuse notation slightly and let $\mathcal{E}$ represent either a set of E-constraints or the equivalence relation they force so we will not use $\sim_{\mathcal{E}}$

etc. unless writing $\mathcal{E}$ would cause confusion. Unless otherwise stated, all equivalence relations are on co-ordinates $\{1, \ldots J\}$. We also do not distinguish between the E-constraint E_{ij} and the set $\{E_{ij}\}$.

Remark It follows that, if $K \in \mathcal{C}_J^d$, in order to calculate $Mxmn(K)$ we should add equivalence relations on $\{x_1, \ldots x_J\}$ which have $J - d$ equivalence classes.

Lemma 110 *If $K \in \mathcal{C}_J^d$, either $d = 0$ or for some j such that $2 \leq j \leq J$, $K' = K + E_{1j}$ is consistent and $K \nRightarrow E_{1j}$.*

Proof Suppose that no such E_{1j} is independent of K. Then for every $j = 2, \ldots J$, $x_1 = x_j$ is either true for every $\vec{x} \in G(K)$ or for no such $\vec{x}$. Hence, for every j, there exists a constant c_j such that $x_j = x_1 + c_j$ for every $\vec{x} \in G(K)$. Now

$$\vec{x} \in G(K) \Rightarrow \vec{x} = (x_1, x_1 + c_2, \ldots, x_1 + c_J) \tag{6.34}$$

for some $x_1 \in \mathbb{R}$. If $d > 0$, then $G(K) = \{(\tau, \tau + c_2, \ldots \tau + c_J) \text{ s.t. } \tau \in \mathbb{R}\}$. However, this contradicts the definition of a constraint set since $V(K)$ would not be bounded. We have proved the lemma.

■

Definition Let $\vec{x} \in \mathbb{R}^J$ and $\mathcal{E}$ be an equivalence relation. We say that $\vec{x}$ *satisfies* $\mathcal{E}$ iff for all i, j such that $i \sim_{\mathcal{E}} j$, $x_i = x_j$.

In the statement and proof of the following algorithm we use the existence of an algorithm $LIST(\vec{x}, m)$ for which the input is $\vec{x} \in \mathbb{R}^J$ together with an integer m and the output, denoted by $Equiv_1, \ldots Equiv_{n(\vec{x})}$, is a listing of the equivalence relations of rank m (i.e. with $J - m$ equivalence classes) that are satisfied by $\vec{x}$.

Algorithm 111 (*Maximin* Calculation Algorithm 2)

Input: *Some* $K \in \mathcal{C}_J^d$.

Output: $Mxmn(K)$.

1 Let $Start = K$ and let K_{dir} be the result of taking the constraints of K and substituting zero for all of the constants on the right hand side of each constraint and substituting the variable u_i for x_i for each $i = 1, \dots J$.[1]

2 For each $j = 2, \dots J$ in turn, add E_{1j} to $Start$ whenever doing so increases $rank(Start)$ and keeps it consistent, stopping when $rank(Start) = J$. Let $\mathcal{E}$ be the set of E-constraints that have been added at this stage. Then let $\vec{x}$ be the unique solution of $Start$.

3 Run $LIST(\vec{x}, d-1)$ to get output $Equiv_1, \dots Equiv_{n(\vec{x})}$. Let $counter = 0$

4 Let i be minimal such that $n(\vec{x}) \geq i > counter$ and $rank(K + Equiv_i) = J - 1$ unless no such i exists, in which case output $\vec{x} = Mxmn(K)$ and stop.

Let $counter = i$, $\mathcal{E}' = Equiv_i$.

5 Find a non-zero solution $\vec{u}$ of $K_{dir} + \mathcal{E}'$. This is the direction parallel to the line $G(K + \mathcal{E}')$. If $\vec{u}$ has sign conflict at lev w.r.t. $\vec{x}$ for some lev, go back to Step 4.

Otherwise, the u_i for which $x_i = lev$ all have the same sign. If that sign is minus, let $\vec{u} = -\vec{u}$.

6 Run the subroutine $Opt(\vec{x}, \vec{u})$ as defined below to obtain a new $\vec{x}$ with new $\mathcal{E}$, then go back to Step 3.

The subroutine $Opt(\vec{x}, \vec{u})$ is given by:

1 For every inc, dec in $\{1, \dots J\}$ s.t. $x_{inc} < x_{dec}$ and $u_{inc} > 0, u_{dec} < 0$ calculate

$$A(inc, dec) = x_{inc} + \frac{u_{inc}(x_{dec} - x_{inc})}{u_{inc} - u_{dec}} \tag{6.35}$$

and define (Inc, Dec) to be the some value of (inc, dec) for which A is minimal.

[1] K_{dir} is a set of equations in variables $u_1, \dots u_J$. It is NOT a constraint set—indeed, if it has a non-negative non-zero solution, we can multiply that solution by large positive numbers to obtain arbitrarily large non-negative solutions.

2 Let new $\mathcal{E} = \mathcal{E}' + E_{Inc,Dec}$, define $\vec{x}$ to be the unique solution of $K + \mathcal{E}$ and stop this subroutine.

Proof The algorithm starts by finding a $\vec{x} \in G(K)$ which is the only solution of K + a set of d independent E_{1j}'s. We can see that this is possible by using Lemma 110 repeatedly. Then the value of $\vec{x}$ is always the only solution of $K + \mathcal{E}$ where $\mathcal{E}$ is some set of d independent E-constraints. At Step 5, we go back to Step 4 iff $\vec{x} = Mxmn(K + \mathcal{E}')$, by Lemma 105.

Otherwise we use the subroutine *Opt* and, by Lemma 112 below, this results in us setting $\vec{x} = Mxmn(K + \mathcal{E}')$. Hence $\vec{x}$ continually improves in the $maximin$ ordering so the algorithm must terminate, say when $\vec{x} = \vec{X}$. Then it remains for us to show that $\vec{X} = Mxmn(K)$.

Lemma 112 *When we run Opt*$(\vec{x}, \vec{u})$ *in Algorithm 111, the new* $\vec{x} = Mxmn(\vec{x} + <\vec{u}>) = Mxmn(K + \mathcal{E}')$ = *the only solution of* $K + (new)\mathcal{E}$.

Proof: We now show that the subroutine *Opt* calculates $Maximin$ of the line $Li = G(K + \mathcal{E}')$. The least x_i for which $u_i \neq 0$ are all such that $u_i > 0$, possibly after using the step "let $\vec{u} = -\vec{u}$". Hence

$$Mxmn(Li) = \vec{x} + \tau\vec{u} \tag{6.36}$$

for some $\tau > 0$, by Lemma 106. By Lemma 105, if $Mxmn(Li) = \vec{y} = \vec{x} + \tau_0\vec{u}$, co-ordinates i s.t. y_i is minimal s.t. $u_i \neq 0$ include j, k such that $u_j u_k < 0$ so in choosing an E-constraint E_{ij} to add to $\mathcal{E}'$ to find $\vec{y}$, we need only consider $E_{inc\,dec}$ for which $x_{inc} < x_{dec}$, $u_{inc} > 0$ and $u_{dec} < 0$. Let A be the common value of co-ordinates inc, dec at the unique point, $\vec{z}$, in Li that satisfies $E_{inc\,dec}$. We will now show that choosing some inc, dec for which this A is minimal will give $\vec{z} = \vec{y}$.

In fact, if co-ordinates up, $down$ are chosen such that $u_{up} > 0, u_{down} < 0$ without imposing any other conditions, then there exists a unique value of τ, say χ, such that $\vec{x} + \chi\vec{u}$ is the point on Li satisfying $E_{up\,down}$ for those $(up, down)$ minimising $A(up, down)$.

Suppose for contradiction that there exist $\tau_1 < \tau_2$ for which the points $\vec{x} + \tau_j\vec{u}$ for $j = 1, 2$ are the unique points on Li satisfying $E_{up_j\,down_j}$ for $j = 1, 2$ such that $A = A_{min} = A(up_j, down_j)$ for $j = 1, 2$.

Now as τ increases from $\tau = \tau_1$, the value at co-ordinate $down_1$ decreases from A and, from the same starting point, x_{up_2} increases past A, so at some $\tau \in (\tau_1, \tau_2)$ there exists a lower meeting value than A using co-ordinates $down_1$ and up_2, so we deduce a contradiction and the required value of τ must be unique.

We now see that $\vec{y} = \vec{x} + \chi\vec{u} = Mxmn(K + \mathcal{E}')$ since $y_{up} = y_{down}$ are minimal among those y_i s.t. $u_i \neq 0$ since if not, a lower y_i s.t. $u_i > 0$ gives a lower meeting value than A (with co-ordinate *down*) at some $\tau > \chi$ and if a lower y_i is s.t. $u_i < 0$ a similar contradiction arises. Hence the lowest meeting value overall is given by $\tau = \chi$ such that $\vec{x} + \chi\vec{u} = Mxmn(Li)$ by Lemma 105. Also $\chi > 0$ so *Opt* finds $Maximin$ by searching those *up* and *down* co-ordinates that meet at a point given by a positive value of τ, (i.e. $x_{up} < x_{down}$). We have proved the lemma.

■

Notation Recall that we have extended the definitions of a loaf, the Blaschke metric and Piecewise Linear Loaf Continuity to constraint sets. Let $K \in \mathcal{C}_J^d$. For all $\vec{c} \in \mathbb{R}^J$ s.t. $\sum_{i=1}^J c_i = 0$, let $K + \vec{c}$ be the result of replacing every occurrence of x_i by $(x_i - c_i)$ in the constraints of K. Note that $\sum_{i=1}^J x_i = 1$ is replaced by $\sum_{i=1}^J (x_i - c_i) = 1$, which is equivalent. Now

$$G(K + \vec{c}) = \{\vec{x} + \vec{c} \text{ s.t. } \vec{x} \in G(K)\} \tag{6.37}$$

and $\{K + \lambda\vec{c}$ s.t. $0 \leq \lambda \leq 1\}$ is a loaf (recall (3.43) for constraint sets) as long as $K + \vec{c}$ is consistent. This should not be confused with $K + K'$ where K' is a constraint set and $+$ means union in that context.

Definition For any $K \in \mathcal{C}_J^d$, let $\vec{x} = K + \mathcal{E}$ be shorthand for $\{\vec{x}\} = G(K + \mathcal{E})$, where $\mathcal{E}$ is an equivalence relation of rank d. In such a case we say that $\vec{x}$ is *visible* (w.r.t. K) and $\mathcal{E}$ is a *sighting* of $\vec{x}$ (w.r.t. K). Let $\vec{x}$ be visible and let $\mathcal{E}'$ be an equivalence relation of rank $d-1$ such that if $K' = K + \mathcal{E}'$, $K' \in \mathcal{C}_J^1$. Then if

$$G(K') = (\vec{x} + <\vec{e'}>) \tag{6.38}$$

we say that $G(K')$ is a *line of sight* (l.o.s.) for $\vec{x}$ (in direction $\vec{e'}$).

The following lemma is helpful for us to show that Algorithm 111 can only terminate at $Mxmn(K)$.

Lemma 113 *For all $K \in \mathcal{C}_J^d$, if $\vec{X}$ is visible w.r.t. K and $\vec{X} = Mxmn(Li)$ for every l.o.s. Li for $\vec{X}$, $\vec{X} = Mxmn(K)$.*

Proof We do this by induction on d.

Base Cases In the case $\underline{d = 0}$, $G(K) = \{\vec{X}\}$, so the lemma is trivial. When $\underline{d = 1}$ and we use $\mathcal{E}'$ to obtain a line of sight for $\vec{x}$, $rank(\mathcal{E}') = 0$ so $\mathcal{E}'$ is the identity equivalence relation. Thus $\vec{X} = Mxmn(K)$.

Inductive Step Assume (I.H.) that the lemma holds for each d such that $d \leq k$, where $1 \leq k < J$. Let K be a fixed consistent constraint set such that $d = k + 1$. Suppose that, for contradiction,

$$Mxmn(K) = \vec{y} \neq \vec{X} \tag{6.39}$$

There must exist an E_{ij} which is independent of K, and is satisfied by $\vec{X}$, such that $y_i < y_j$, say. Let $K_{aug} = K + E_{ij}$. Then $dim(G(K_{aug})) = k$ and by I.H. $\vec{X} = Mxmn(K_{aug})$ since otherwise there would be an equivalence relation $\mathcal{F}$ such that $G(K_{aug} + \mathcal{F})$ is a l.o.s. for $\vec{X}$ such that $\vec{X} \neq Mxmn(K_{aug} + \mathcal{F})$ so we would have found a contradiction.

Consider the loaf $\hat{K}_{aug}$ which has slices given by

$$K_{aug}(\lambda) = K_{aug} + \lambda\vec{c} \tag{6.40}$$

and λ-range $[0, 1]$, where $\vec{c} = \vec{y} - \vec{X}$. For all $\lambda \in [0, 1]$, let

$$\vec{z}(\lambda) = \vec{X} + \lambda(\vec{y} - \vec{X}) \in G(K_{aug}(\lambda)) \subset G(K) \tag{6.41}$$

Since $\vec{y} = Mxmn(K)$, $\vec{z}(\lambda)$ is *maximin*-better than $\vec{z}(\mu)$ iff $1 \geq \lambda > \mu \geq 0$, by Lemma 106, so $\vec{Y}(\lambda) = Mxmn(K_{aug}(\lambda))$ is *maximin*-better than $\vec{X}$ for all $\lambda \in (0, 1]$.

For all $\lambda \in [0,1]$, $\vec{Y}(\lambda)$ can be expressed in the form $\vec{Y} = K_{aug}(\lambda) + \mathcal{E}$ for some non-trivial equivalence relation $\mathcal{E}$ such that $rank(\mathcal{E}) = k$ and $K \not\Rightarrow E_{vw}$ for each $E_{vw} \in \mathcal{E}$. Now Piecewise Linear Loaf Continuity of $Maximin$ holds for constraint sets (due to Theorem 88 for constraint sets), so for small enough positive values of ϵ

$$Y_{v'}(\epsilon) = Y_{w'}(\epsilon) \Rightarrow X_{v'} = X_{w'} \tag{6.42}$$

for such a $(v,w) = (v',w')$.

Since $K \not\Rightarrow E_{v'w'}$, we can see that $\vec{X} = Mxmn(K + E_{v'w'})$ by a similar argument to that used to show that $\vec{X} = Mxmn(K_{aug})$, but $\vec{Y}(\epsilon) \in G(K + E_{v'w'})$ is *maximin*-better than $\vec{X}$, giving a contradiction. We have proved the lemma.

■

Completing the proof of Algorithm 111 Recall that the algorithm terminates at $\vec{x} = \vec{X}$. This occurs because we have used Steps 4 and 5 to check each line of sight for $\vec{X}$ which is the solution set of $K + Equiv_i$ and have found, in every case, that $\vec{X} = Mxmn(K + Equiv_i)$. By Lemma 113, $\vec{X} = Mxmn(K)$ and we have proved the algorithm.

■

Remark This algorithm can still seem a little disappointing. Do we really have to check every l.o.s. for $\vec{X}$ to be sure that $\vec{X} = Mxmn(K)$? Although we can refine the above, it is certainly NOT enough just to check d linearly independent such lines, as the following example shows:

$$K = \left\{ \sum_{i=1}^{8} x_i = 1,\ x_7 = 0,\ x_8 = 0,\ x_1 = \frac{5}{12} - x_4,\ x_2 = \frac{7}{12} - 2x_4,\ x_5 = x_3 + x_4 - 1/4 \right\} \tag{6.43}$$

and $\vec{X} = \frac{1}{12}(3,3,2,2,1,1,0,0)$.

In this example, $d = dim(G(K)) = 2$ and $\{\vec{X}\} = G(K + E_{12} + E_{34})$. $S_{12} = G(K + E_{12})$ and $S_{34} = G(K + E_{34})$ are two l.o.s. for $\vec{X}$ which span $G(K)$.

Indeed $(0,0,1,0,1,-2,0,0)$ is parallel to S_{12} and the co-ordinates of lowest value in $\vec{X}$ which have non-zero rate of change along S_{12} are $X_5 = X_6 = 1/12$ and their change is of opposite sign so $Mxmn(K+E_{12}) = \vec{X}$. Similarly $(1,2,-1,-1,-2,1,0,0)$

is parallel to S_{34} and similar inspection shows that $\vec{X} = (Mxmn(K + E_{34}))$. However $S_{56} = G(K + E_{56})$ is a l.o.s. for $\vec{X}$, in direction $(1, 2, 0, -1, -1, -1, 0, 0)$, and

$$\frac{1}{36}(7, 5, 6, 8, 5, 5, 0, 0) = Mxmn(K + E_{56}) \neq \vec{X} \tag{6.44}$$

Intuitively we could believe that checking that $\vec{X} = Mxmn(K + E_{56})$ seems more important because $X_5 = X_6 < X_3 = X_4 < X_1 = X_2$ so E_{56} is the "minimal non-trivial E-constraint satisfied by $\vec{X}$", an idea that will now be formalised.

Notation If $\mathcal{E}'$, $\mathcal{E}$ are equivalence relations such that $rank(\mathcal{E}') = d - 1$, $rank(\mathcal{E}) = d$ and

$$i \sim_{\mathcal{E}} j \Longrightarrow i \sim_{\mathcal{E}'} j \tag{6.45}$$

for all co-ordinates i, j, we write $\mathcal{E}' <_1 \mathcal{E}$. Thus $\mathcal{E}$ may be thought of as being obtained by gluing together a certain pair of equivalence classes of $\mathcal{E}'$. If, for a fixed sighting $\mathcal{E}$ of $\vec{x}$ w.r.t. $K \in \mathcal{C}_J^d$, $\vec{x} = Mxmn(K + \mathcal{E}')$ for every $\mathcal{E}'$ such that $\mathcal{E}' <_1 \mathcal{E}$, we say that $\vec{x}$ is *Pseudo-Maximin* of K w.r.t. $\mathcal{E}$, abbreviated to $PsMxmn(K)$ w.r.t. $\mathcal{E}$.

Remark We can see, from the example of (6.43), that if $\vec{x}$ is $PsMxmn(K)$ w.r.t. $\mathcal{E}$, then it is not necessarily true that $\vec{x} = Mxmn(K)$. Nevertheless, as we show in Theorem 114 below, there is an important connection between the concepts *Maximin* and *Pseudo-Maximin*.

Notation If $\vec{x}$ is visible w.r.t. $K \in \mathcal{C}_J^d$ and e is the least value appearing in $\vec{x}$ such that there exist i, j s.t. $x_i = x_j = e$ and $K \not\Rightarrow E_{ij}$, then those E_{ij} are the *principal equalities* of $\vec{x}$ w.r.t. K.

The *principal sightings* of a (visible) $\vec{x}$ w.r.t. K are defined recursively w.r.t. d as follows. If $d = 1$ a principal sighting of $\vec{x}$ w.r.t. K is a principal equality of $\vec{x}$ w.r.t. K.

In general a sighting $\mathcal{E}$ of $\vec{x}$ w.r.t. $K \in \mathcal{C}_J^d$ is *principal* iff it can be expressed in the form $\mathcal{E} = \mathcal{E}' + E_{ij}$ ($\mathcal{E}' <_1 \mathcal{E}$), where E_{ij} is principal and $\mathcal{E}'$ is a principal sighting of $\vec{x}$ w.r.t. $(K + E_{ij}) \in \mathcal{C}_J^{d-1}$.

In other words, choose a "minimal non-trivial E-constraint E_{ij} satisfied by $\vec{x}$" and add that to K before choosing another minimal non-trivial equality (w.r.t. $K + E_{ij}$), and so on, and this process will stop with a principal sighting.

We use these concepts throughout this chapter.

Theorem 114 *Let $K \in \mathcal{C}_J^d$. The following are equivalent:*

(i) $\vec{x} = Mxmn(K)$.

(ii) $\vec{x}$ is visible w.r.t. K and for every principal sighting $\mathcal{E}$ of $\vec{x}$ w.r.t. K, $\vec{x} = PsMxmn(K)$ w.r.t. $\mathcal{E}$.

Proof$(i) \Rightarrow (ii)$ This follows directly from the definition of $PsMxmn$.

Remark For the example K given in (6.43) we can see that $E_{12}+E_{34}$, $E_{12}+ E_{56}$, $E_{34} + E_{56}$ are the sightings of $\vec{x}$, of which only the last one is principal so, by Theorem 114, we need only check whether $\vec{X} = Mxmn(K + E_{34})$ and $\vec{X} = Mxmn(K + E_{56})$. Since $\vec{X} \neq Mxmn(K + E_{56})$, we can see a little quicker that $\vec{X} \neq Mxmn(K)$, by not checking that $\vec{X} = Mxmn(K + E_{12})$.

Thus, in general, sometimes the theorem above may make it easier for us to check whether $\vec{x} = Mxmn(K)$.

Notation This notation applies to rest of this subsection. We suppose that $K \in \mathcal{C}_J^d$, where $d > 0$. Then, if $\vec{x} = \vec{X}$ is visible w.r.t. K, a finite sequence of consistent constraint sets $K_1, K_2 \ldots K_p$, a sequence of real numbers $e_1, e_2, \ldots e_p$ and of integers $r_1, r_2, \ldots r_p$ are defined as follows. By convention, we define $K_0 = K$ and $r_0 = 0$.

Let e_1 be minimal such that for some co-ordinates i, j, $X_i = X_j = e_1$ and $K \not\Rightarrow E_{ij}$. Then let

$$K_1 = K + \{E_{ij} \,|\, X_i = X_j = e_1\} \tag{6.46}$$

and let $rank(K_1) = J - d + r_1$, where $r_1 > 0$. If K_k has been defined s.t. $rank(K_k) = J - d + r_k \neq J$ we let e_{k+1} be minimal such that for some i, j, $X_i = X_j = e_{k+1}$ and $K_k \not\Rightarrow E_{ij}$, then

$$K_{k+1} = K_k + \{E_{ij} \,|\, X_i = X_j = e_{k+1}\} \tag{6.47}$$

By considering $\vec{X}$ as $K + \mathcal{E}$ for some sighting $\mathcal{E}$, if $rank(K_k) < J$, the E-constraints added so far to K do not, in the presence of K, gen-imply $\mathcal{E}$; so further E-constraints can continue to be added until some K_k is s.t. $rank(K_k) = J$. In this case, we define $p = k$. Then $\{\vec{X}\} = G(K_p)$ and

$$r_p = d > r_{p-1} > \ldots > r_1 \tag{6.48}$$

If $d = 0$, a sighting of $\vec{x}$ w.r.t. K must be the trivial equivalence relation—in this case $p = 0$ and there are no e_i's, r_i's or K_i's defined. This concludes the definitions of the e_i's, r_i's, K_i's and p, which depend on K and $\vec{X}$; we suppress that dependence wherever this does not cause confusion.

The following lemma helps us to characterise the principal sightings of $\vec{X}$ w.r.t. $K \in \mathcal{C}_J^d$ and ultimately to prove Theorem 114 in the direction $((ii) \Rightarrow (i))$.

Lemma 115 *Let $\vec{x} \in \mathbb{R}^J$ be visible w.r.t. $K \in \mathcal{C}_J^d$. A sighting $\mathcal{E}$ of $\vec{x}$ w.r.t. K, written as a list of d E-constraints, is principal iff, for each $g = 1, \ldots p$, $r_g - r_{g-1}$ of them are of the form E_{ij} s.t. $x_i = x_j = e_g$.*

Proof $(\Rightarrow)$ Fix a principal sighting $\mathcal{E}$ of such an $\vec{x}$ w.r.t. K, say $\vec{x} = \vec{X}$. By definition of a principal sighting, we fix a sequence $E_{i_1 j_1}, E_{i_2 j_2}, \ldots E_{i_d j_d}$ for which $E_{i_{k+1} j_{k+1}}$ is a principal equality of $\vec{X}$ w.r.t. $K + E_{i_1 j_1} + \ldots + E_{i_k j_k}$ for each $k = 0, \ldots d-1$. We show that

$$X_{i_1} = \ldots = X_{i_{r_1}} = e_1,\ X_{i_{(r_1+1)}} = \ldots = X_{i_{r_2}} = e_2,\ \ldots X_{i_{(r_u)}} = e_u \tag{6.49}$$

by induction on u, which varies from 0 up to p.

Base Case $\underline{u = 0}$ In this case the lemma is trivial.

Inductive Step Assume (I.H.) that the result holds for $u = k < p$. Then, by definition of a principal sighting, the next E-constraint, $E_{i_{(r_k+1)} j_{(r_k+1)}}$, is a principal equality of $\vec{X}$ w.r.t.

$$K_{(k)} = K + E_{i_1 j_1} + \ldots + E_{i_{r_k} j_{r_k}} \tag{6.50}$$

Claim $K_{(k)} \equiv_G K_k$.

Proof of claim We observe that if $\vec{x} \in G(K_k)$, every pair of co-ordinates which are equal in value in $\vec{X}$ with common value less than or equal to e_k are equal in $\vec{x}$, so

$$\vec{x} \in G(K_k) \Rightarrow \vec{x} \in G(K_{(k)}) \tag{6.51}$$

However by the Inductive Hypothesis,

$$rank(K_k) = rank(K_{(k)}) = J - d + r_k \tag{6.52}$$

forcing K_k and $K_{(k)}$ to be gen-equivalent and we have proved the claim.

Suppose that, for contradiction, v is a positive integer such that $v \leq r_{k+1} - r_k$, and is minimal such that $X_{i_{(r_k+v)}} \neq e_{k+1}$.

Let

$$K_{(k)}^v = K_{(k)} + E_{i_{(r_k+1)}j_{(r_k+1)}} + \ldots + E_{i_{(r_k+v)}j_{(r_k+v)}} \tag{6.53}$$

If $X_i = X_j = c$ and $c < e_{k+1}$ then $K_k \Rightarrow E_{ij}$, so $K_{(k)}^v \Rightarrow E_{ij}$ and hence $X_{i_{(r_k+v)}} \geq e_{k+1}$. Now $v \leq r_{k+1} - r_k$ and $X_{i_{(r_k+v)}} > e_{k+1}$ so

$$K_{(k)} + E_{i_{(r_k+1)}j_{(r_k+1)}} \ldots + E_{i_{(r_k+v-1)}j_{(r_k+v-1)}} \Rightarrow \{E_{ij} \mid X_i = X_j = e_{k+1}\} \tag{6.54}$$

and, consequently, the rank of the LHS is at least r_{k+1}, which is a contradiction as there are not enough constraints. Hence we deduce that no such v exists. This completes the Inductive Step and the proof of ($\Rightarrow$).

($\Leftarrow$): We express a sighting $\mathcal{E}$ as a sequence of E_{ij}'s where the first r_1 are of the form E_{ij}'s s.t. $X_i = X_j = e_1$ and the next $r_2 - r_1$ are of the form E_{ij}'s s.t. $X_i = X_j = e_2$, etc.. $E_{i_1 j_1}$ is a principal equality of $\vec{X}$ w.r.t. K so it is sufficient to show that each $E_{i_{(k+1)}j_{(k+1)}}$ is a principal equality of $\vec{X}$ w.r.t. $K' = K + E_{i_1 j_1} \ldots + E_{i_k j_k}$. We consider two cases:

Case 1 $\underline{X_{i_{(k+1)}} = X_{i_k} (= \text{some } e_q)}$ Suppose for contradiction that $E_{i_{(k+1)}j_{(k+1)}}$ is not a principal equality of $\vec{X}$ w.r.t. K'. Then there exist i, j s.t. $X_i = X_j = c < e_q$ and $K' \not\Rightarrow E_{ij}$ so $K* = K + E_{i_1 j_1} + \ldots + E_{i_{(k-1)}j_{(k-1)}} \not\Rightarrow E_{ij}$. However then $E_{i_k j_k}$ is not a principal equality of $\vec{X}$ w.r.t. $K*$, contradicting the defining property of $E_{i_k j_k}$. Hence we have proved ($\Leftarrow$) in this case.

Case 2 $\underline{X_{i_k} = e_q,\ X_{i_{(k+1)}} = e_{q+1}}$. Since $K' = K+$ E-constraints of the form E_{ij} for which $X_i = e_1, e_2, \ldots$ or e_q and $rank(K') = J - d + r_q$ then $K' \equiv_G K_q$ as in (6.52) and the principal equalities of $\vec{X}$ w.r.t. K' are all of the form E_{ij} s.t. $X_i = e_{q+1}$ by the definition of e_{q+1}. We have proved ($\Leftarrow$) in this case.

We have proved the lemma.

■

Notation In the following proofs, we shall write expressions of the form $\vec{a} + \sum_{i=1}^{d} <\vec{u}^{(i)}>$ as shorthand for $(\vec{a} + <\vec{u}^{(1)}, \ldots \vec{u}^{(d)}>)$.

The following theorem, which helps us prove Theorem 114, is an exercise in linear algebra.

Theorem 116 *Suppose that* $S \in \mathcal{C}_J^g$. *Assume that* $A_1, A_2, \ldots A_g$ *are linearly independent constraints such that* $S + \{A_1, \ldots A_g\}$ *has a unique solution* $\vec{x}$. *For* $i = 1, \ldots g$ *we define the constraint set* $Aug_i(S)$ *by* $S + \{A_j \,|\, j \neq i\}$ *and let* $\vec{b}^{(i)}$ *be fixed such that*

$$G(Aug_i(S)) = \vec{x} + <\vec{b}^{(i)}> \tag{6.55}$$

Then $G(S) = \vec{x} + \sum_{i=1}^{g} <\vec{b}^{(i)}>$.

Proof We do this by induction on g, letting the unique solution of the constraint set $S + \{A_1, \ldots A_g\}$ be given by $\vec{x} = \vec{X}$.

Base Case $\underline{g = 1}$ In this case, $S + A_1$ has a unique solution $\vec{X}$ and the solution set of $Aug_1(S) = S$ is $\vec{X} + <\vec{b_1}>$ so in this case the theorem is trivial.

Inductive Step Assume (I.H.) that the theorem holds when $g \leq k$ and suppose that the hypotheses of the theorem apply for the case $g = k + 1$. Then let $S* = S + A_{k+1} \in \mathcal{C}_J^k$, since $S \not\Rightarrow A_{k+1}$. For each $i = 1, \ldots k$, we now define

$$Aug_i(S*) = S* + \{A_j \,|\, 1 \leq j \leq k \ \text{ and } \ j \neq i\} \tag{6.56}$$

Now $Aug_1(S) \equiv_G Aug_1(S*), \ldots Aug_k(S) \equiv_G Aug_k(S*)$ have solution sets $\vec{X} + <\vec{b}^{(i)}>$ respectively and by the I.H.

$$G(S*) = \vec{X} + \sum_{i=1}^{k} <b^{\vec{(i)}}> \tag{6.57}$$

which is a k-dimensional affine subset of $G(S)$. Suppose for contradiction that every generalised solution of $Aug_{k+1}(S)$ satisfies $S*$. Then $Aug_{k+1}(S) \Rightarrow A_{k+1}$ so $S + A_1 + \ldots + A_k \Rightarrow A_{k+1}$, which is a contradiction.

Hence the line $\vec{X} + <\vec{b}^{(k+1)}>$ does not lie in $\vec{X} + \sum_{i=1}^{k} <\vec{b}^{(i)}>$ and so $\vec{X} + \sum_{i=1}^{k+1} <\vec{b}^{(i)}>$ is a $(k+1)$-dimensional affine subset of $G(S)$; hence it must be $G(S)$, completing the Inductive Step and the proof of the theorem.

■

Remark In particular, Theorem 116 implies that if $\mathcal{E}* = E_{i_1j_1} + E_{i_2j_2} \ldots + E_{i_dj_d}$ is a specific principal sighting of $\vec{x}$ w.r.t. K as in the hypotheses of Theorem 114, the lines of sight Li_q given by

$$Li_q = G(K + \mathcal{E}*\backslash E_{i_qj_q}) = \vec{x} + <\vec{dir}^{(q)}> \tag{6.58}$$

for $q = 1, \ldots d$, produce the set $\{\vec{dir}_i | 1 \leq i \leq d\}$ as a basis for $D(K)$ since $G(K) = \vec{x} + \sum_{i=1}^{d} <\vec{dir}_i>$.

The lemma that follows helps us to see how to identify principal sightings.

Lemma 117 *Let E_{ij} be an E-constraint appearing in at least one principal sighting of $\vec{x}$ w.r.t. $K \in \mathcal{C}_J^d$, where $d > 0$. Then $\vec{x}$ is visible w.r.t. $K* = K + E_{ij}$, and every principal sighting $\mathcal{E}'$ of $\vec{x}$ w.r.t. $K*$ is such that $\mathcal{E}' + E_{ij}$ is a principal sighting of $\vec{x}$ w.r.t K.*

Proof We require the following sublemma.

Sublemma 118 *Let $E_{i'j'}$ be an E-constraint such that $E_{i'j'} \in \mathcal{E}_0$, where $\mathcal{E}_0$ is a principal sighting of $\vec{X}$ w.r.t. $K \in \mathcal{C}_J^d$, and $d > 0$. Then $\vec{X}$ is visible w.r.t. $K + E_{i'j'}$ and $\mathcal{E}_0 \backslash \{E_{i'j'}\}$ is a principal sighting of $\vec{X}$ w.r.t. $K + E_{i'j'}$.*

Proof Assume the hypotheses of the sublemma. Let $K' = K + E_{i'j'}$. It is trivial that $\vec{X}$ is visible w.r.t. K' by observing that $\vec{X} = K + \mathcal{E}_0 = K + E_{i'j'} + \mathcal{E}_0 \backslash \{E_{i'j'}\}$.

Now we fix a way of writing $\mathcal{E}_0$ as a sequence $\{E_{i_1j_1}, E_{i_2j_2}, \ldots E_{i_dj_d}\}$, such that $E_{i_1j_1}$ is a principal equality of $\vec{X}$ w.r.t. K and, for each $t = 2, \ldots d$, $E_{i_tj_t}$ is a principal equality of $\vec{X}$ w.r.t. $K + \{E_{i_1j_1}, \ldots E_{i_{(t-1)}j_{(t-1)}}\}$. Suppose that $E_{i'j'} = E_{i_qj_q}$.

Claim:

- For all $t = 1, \ldots q-1$, $E_{i_t j_t}$ is a principal equality of $\vec{X}$ w.r.t.

  $K' + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}\}$.

- For all $t = q+1, \ldots d$, $E_{i_t j_t}$ is a principal equality of $\vec{X}$ w.r.t.

  $K' + \{E_{i_1 j_1}, \ldots E_{i_{(q-1)} j_{(q-1)}}, E_{i_{(q+1)} j_{(q+1)}}, \ldots E_{i_{(t-1)} j_{(t-1)}}\}$.

Proof of claim: Suppose that t is an integer less than q. Since $rank(K + \mathcal{E}_0) = J$, $rank(K) = J - d$ and there are d $E_{i_k j_k}$'s, $K' + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}\} \equiv_G K + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}, E_{i_q j_q}\} \not\Rightarrow E_{i_t j_t}$.

Suppose for contradiction that there exists E_{ij} which is satisfied by $\vec{X}$, and that $X_i = X_j < X_{i_t}$, and $K' + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}\} \not\Rightarrow E_{ij}$. Then

$$K + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}\} \not\Rightarrow E_{ij} \tag{6.59}$$

so $E_{i_t j_t}$ is not a principal equality of $\vec{X}$ w.r.t. $K + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}\}$ and we have reached a contradiction, proving the first part of the claim.

For the second part of the claim,

$$K' + \{E_{i_1 j_1}, \ldots E_{i_{(q-1)} j_{(q-1)}}, E_{i_{(q+1)} j_{(q+1)}}, \ldots E_{i_{(t-1)} j_{(t-1)}}\} \equiv_G K + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}\} \tag{6.60}$$

and $E_{i_t j_t}$ is a principal equality of $\vec{X}$ w.r.t. $K + \{E_{i_1 j_1}, \ldots E_{i_{(t-1)} j_{(t-1)}}\}$. We have proved the claim.

■

Proof of Sublemma 118 continued: Using the claim and the definition of a principal sighting, we deduce that $\mathcal{E}_0 \backslash \{E_{i'j'}\}$ is a principal sighting of $\vec{X}$ w.r.t. K' and we have proved the sublemma.

■

Proof of Lemma 117 continued: Let $\vec{x} = \vec{X}$ be visible w.r.t. $K \in \mathcal{C}_J^d$, where $d > 0$. Let $E_{i'j'}$ be a specific E-constraint which appears in a principal sighting, $\mathcal{E}_0$, of $\vec{X}$ w.r.t. K. Suppose that $K' = K + E_{i'j'}$.

By Sublemma 118, $\mathcal{E}_0 \backslash \{E_{i'j'}\}$ is a principal sighting of $\vec{X}$ w.r.t. K'.

Let $\mathcal{E}$ be some principal sighting of $\vec{X}$ w.r.t. K'. Then, by Lemma 115, when we write $\mathcal{E}$ or $\mathcal{E}_0 \backslash \{E_{i'j'}\}$ as a list of $d-1$ E-constraints we find that, for each $g = 1, \ldots p(K')$, $r_g - r_{g-1}$ of them are of the form E_{ij} s.t. $x_i = x_j = e_g(K')$. Hence, for all $e \in \mathbb{R}$, $\mathcal{E} + E_{i'j'}$ and $\mathcal{E}_0$ contain the same number of E-constraints E_{ij} such that $X_i = X_j = e$. Since $\mathcal{E}_0$ is a principal sighting of $\vec{X}$ w.r.t. K, we use Lemma 115 again to deduce that $\mathcal{E} + E_{i'j'}$ is also a principal sighting of $\vec{X}$ w.r.t. K. We have proved the lemma.

■

Lemma 119 *Let $\vec{X}$ be visible w.r.t. $K \in \mathcal{C}_J^d$, such that $\vec{X} = PsMxmn(K)$ w.r.t. every principal sighting of $\vec{X}$ w.r.t. K. Also assume that for all $K' \in \mathcal{C}_J^{d'}$ s.t. $d' < d$, and each $\vec{x}$ visible w.r.t. K', Theorem 114 is true.*

Let $\mathcal{E}$ be a fixed principal sighting of $\vec{X}$ w.r.t. K. In the manner of (6.58) *let $Aug_q = K + \mathcal{E} \backslash E_{i_q j_q}$ for each $q = 1, \ldots d$ and let $Li_1, \ldots Li_d$ be the lines of sight given by*

$$Li_q = G(Aug_q) = \vec{x} + <\vec{dir}^{(q)}> \tag{6.61}$$

so that, by Theorem 116, the set $\{\vec{dir}^{(1)}, \ldots \vec{dir}^{(d)}\}$ is a basis for $D(K)$.

Then for all E_{ij} which appear in a principal sighting of $\vec{X}$ w.r.t. K, $\vec{X} = Mxmn(K + E_{ij})$.

Also, if q' is such that $r_{m-1} < q' \leq r_m$, then $dir_i^{(q')} = 0$ for all i s.t. $X_i < e_m$, and of the values of $dir_i^{(q')}$ for which $X_i = e_m$, some are positive and some are negative.

Proof Since $\vec{X} = PsMxmn(K)$ w.r.t. every principal sighting of $\vec{X}$ w.r.t. K, we use Lemma 117 to deduce that for each E_{ij} which appears in a principal sighting of $\vec{X}$ w.r.t. K, $\vec{X} = PsMxmn(K')$ w.r.t. every principal sighting of $\vec{X}$ w.r.t. K', where $K' = K + E_{ij}$. For every such K', $dim(G(K')) < dim(G(K))$, so Theorem 114 holds for these K' and $\vec{X} = Mxmn(K + E_{ij})$ for every E_{ij} which appears in a principal sighting of $\vec{X}$ w.r.t. K.

Since $Maximin$ satisfies Obstinacy (Theorem 72 for constraint sets), then for every $q = 1, \ldots d$,

$$\vec{X} = Mxmn(Aug_q) = Mxmn(\vec{X} + <\vec{dir}^{(q)}>) \tag{6.62}$$

Using $q = q'$ as in the statement of the lemma, $Aug_{q'} \Rightarrow K_{m-1}$ but $Aug_{q'} \not\Rightarrow K_m$ (using Lemma 115) so e_m is minimal such that there exists an E-constraint $E_{i'j'}$ such that $X_{i'} = X_{j'} = e_m$ but $L_{q'} \not\Rightarrow E_{i'j'}$. By Lemma 105, $\vec{dir}^{(q')}$ has sign conflict at some lev and the following claim is now sufficient to prove the lemma.

Claim $lev = e_m$.

Proof of claim If $lev < e_m$, there exist $i*, j*$ s.t.

$$X_{i*} = X_{j*} = lev \text{ and } dir^{(q')}_{i*} dir^{(q')}_{j*} < 0 \tag{6.63}$$

so $dir^{(q')}_{i*} \neq dir^{(q')}_{q*}$ and $Aug_{q'} \not\Rightarrow E_{i*j*}$ although E_{i*j*} is satisfied by $\vec{X}$, giving a contradiction.

If $lev > e_m$, $dir^{(q')}_i = 0$ for each i s.t. $X_i \leq e_m$ but in that case $Aug_q \Rightarrow K_m$, also causing a contradiction. Hence $lev = e_m$ and we have proved the claim.

We have proved the lemma.

■

In the final run-up to proving Theorem 114, we assume for contradiction that $\vec{X} \neq Mxmn(K)$.

Lemma 120 *Assume that K, $\vec{X}$ are as in the first paragraph of the statement of Lemma 119. We also assume that $\vec{X} \neq Mxmn(K)$. Then there exists a non-zero direction $\vec{z} \in D(K)$ s.t. $\vec{X} \neq Mxmn(\vec{X} + <\vec{z}>)$ and the least y for which there exists i s.t. $X_i = y$ and $z_i \neq 0$ is e_1.*

Proof Let $\vec{z} = (Mxmn(K) - \vec{X}) \neq \vec{0} \in D(K)$, then by Obstinacy of $Maximin$ (Theorem 72 for constraint sets),

$$Mxmn(K) = Mxmn(\vec{X} + <\vec{z}>) \tag{6.64}$$

Now suppose for contradiction that the least y for which there exists i s.t. $x_i = y$ and z_i non-zero, say $y = Y$, is greater than e_1. Then for any principal equality E_{ij} of $\vec{X}$ w.r.t. K,

$$X_i = X_j = e_1 \text{ so } z_i = z_j = 0 \tag{6.65}$$

and the line $(\vec{X} + <\vec{z}>)$ lies in $G(K + E_{ij})$ but then by Obstinacy

$$Mxmn(K) = Mxmn(K + E_{ij}) \text{ and } \vec{X} \neq Mxmn(K + E_{ij}) \tag{6.66}$$

contradicting Lemma 119.

Now we suppose for contradiction that $Y < e_1$. We know that $\vec{z} \in \sum_{q=1}^{d} <\vec{dir}^{(q)}>$ and for each i s.t. $X_i < e_1$ and every $q = 1, \ldots d$, $dir_i^{(q)} = 0$ so we have reached a contradiction.

Hence $Y = e_1$ and we have proved the lemma.

■

Proof of Theorem 114 $(ii) \Rightarrow (i)$ We continue to assume the conditions of Lemmas 119 and 120 including the assumption, for contradiction, that
$\vec{X} \neq Mxmn(K)$.

Since $\vec{X} \neq Mxmn(\vec{X} + <\vec{z}>)$ we use Lemma 105 and Lemma 120 to see that for all $i \in C$, where C is given by

$$C = \{i \,|\, 1 \leq i \leq J \text{ and } X_i = e_1\} \tag{6.67}$$

the non-zero z_i have the same sign, and w.l.o.g. these values of z_i are all non-negative as replacing $\vec{z}$ by $-\vec{z}$ does not prevent it satisfying the conditions required in the statement of Lemma 120.

There exists an integer $q' = 1, 2, \ldots$ or d and $k \in C$ such that $dir_k^{(q')} \neq 0$ and $z_k > 0$ since $\vec{z}$ is a linear combination of the $\vec{dir}^{(q)}$'s. By Lemma 119, $dir_i^{(q')} = 0$ for all i such that $X_i < e_1$ and, of those $dir_i^{(q')}$ for which $i \in C$, some values are positive and others negative. Consider the points (all in $D(K)$) given by

$$push(\tau) = \vec{z} + \tau\vec{dir}^{(q')} \tag{6.68}$$

where $\tau \in \mathbb{R}$. For any τ and any i such that $X_i < e_1$, $push(\tau)_i = 0$, by Lemma 119. Since for every $i \in C$, the z_i are non-negative and not all zero and there exist $i, j \in C$ such that $dir_i^{(q')} dir_j^{(q')} < 0$, we deduce that for all $\tau \in \mathbb{R}$ there exists $i \in C$ s.t. $push(\tau)_i \neq 0$.

Thus when τ is fixed it is always possible, using Lemma 105, to see whether

$$\vec{X} = Mxmn(\vec{X} + <push(\tau)>) \tag{6.69}$$

by looking for whether there are both positive and negative $push(\tau)_i$ for the $i \in C$.

There exists a fixed real interval $[v, w]$ s.t. $v \leq 0 \leq w$ and

$$push(\tau)_i \geq 0 \text{ for all } i \in C \tag{6.70}$$

if and only if $\tau \in [v, w]$. Hence for some $i_1 \in C$, $push(v)_{i_1} = 0$ and $dir_{i_1}^{(q')} > 0$ and for some $i_2 \in C$, $push(w)_{i_2} = 0$ and $dir_{i_2}^{(q')} < 0$. Hence there exists

$$\tau_0 \in [v, w] \text{ s.t. } push(\tau_0)_{i_1} = push(\tau)_{i_2} \text{ but } \vec{X} \neq Mxmn(\vec{X} + <push(\tau_0)>) \tag{6.71}$$

and $E_{i_1 i_2}$ is a principal equality of K, contradicting the fact that $\vec{X} = Mxmn(K + E_{i_1 i_2})$ from Lemma 119. Hence our assumptions are contradictory and we have proved the theorem.

■

6.5 *Maximin* Calculation Algorithm 3

Algorithm 121 (*Maximin* **Calculation Algorithm 3)**

Input: *Some* $K \in \mathcal{C}^d_J$.

Output: $Mxmn(K)$.

1 Let K_{dir} be the result of taking the constraints of K and substituting zero for all of the constants on the right hand side of each constraint and substituting the variable u_i for x_i for each $i = 1, \ldots J$.

2 Fix an initial value for $\vec{x} \in G(K)$ by adding E-constraints of the form E_{1j} to K until a system $K*$ of rank J is reached, just as in Step 2 of MCA 2. Let $\vec{x}$ be the unique point in $G(K*)$. If $d = 0$, output $\vec{x}$ and stop.

3 Add $d-1$ equations of the form $u_i = 0$ to K_{dir} in every possible way and for those resulting systems with rank $J-1$ retain a non-zero solution and its negation. These vectors are the members of V, given by $V = \{\pm\vec{v}^{(1)}, \ldots \pm \vec{v}^{(s)}\}$. Let $stable = \{i = 1 \ldots J$ s.t. $v_i = 0$ for all $i \in V\}$.

4 Let $minmobile =$ $\{i = 1, \ldots\ J$ not in $stable$ such that $x_i = min(\{x_i$ s.t. $i \notin stable)\}\}$.

5 Let $\vec{trydir}$ be the sum of those $\vec{v} \in V$ such that $v_i \geq 0$ for all $i \in minmobile$.

6 Let $newstable$ be the set of those $i \in minmobile$ such that $trydir_i = 0$. If $newstable = \emptyset$, go to Step 9.

7 Delete from V all $\vec{v} \in V$ for which, for some i, $v_i \neq 0$ and $i \in newstable$; then let $stable = stable \cup newstable$. If V is empty, output $\vec{x}$ and stop. If $newstable = minmobile$, go to Step 4.

8 Let $minmobile = \{i = 1 \ldots J$ not in $stable$ such that $x_i = min(\{x_i$ s.t. $i \notin stable)\}$.

9 Let $K_{try} = K_{dir} \cup \{(u_i = 0) \text{ for every } i \in stable\} \cup \{(u_i = 1) \text{ for every } i \in minmobile\}$. If K_{try} has real solutions, let $\vec{U}$ denote a specific solution, and set $\vec{trydir} = \vec{U}$.

10 Fix some $\vec{y} = \vec{x} + \lambda \vec{trydir}$ such that λ is positive and is minimal such that one of the co-ordinates in *minmobile* is equal to some co-ordinate j not in *stable*. Let $\vec{x} = \vec{y}$ and go to Step 4.

Remark Step 10 does not necessarily do $\vec{x} = Mxmn(\vec{x} + < \vec{trydir} >)$ but it does make $\vec{x}$ *maximin*−better. Note that $D(K)$ is the set of solutions of the equations K_{dir}.

To prove the algorithm works, we first clarify what the initial list V is for:

Lemma 122 *Let K be the input for $MCA3$ and let the initial value of V be V_0. Then for each disjoint $S, T \subseteq \{1, 2, \ldots J\}$, each direction $\vec{w} \in D(K)$ for which $w_i \geq 0$ for each $i \in S$ and $w_j = 0$ for all $j \in T$ is of the form*

$$\lambda_1 \vec{v}^{(i_1)} + \lambda_2 \vec{v}^{(i_2)} + \ldots + \lambda_n \vec{v}^{(i_n)} \tag{6.72}$$

where $\vec{v}^{(i_1)} \ldots \vec{v}^{(i_n)}$ lists the $\vec{v} \in V_0$ such that $v_i \geq 0$ for every $i \in S$ and $v_j = 0$ for all $j \in T$ and all of the λ_j are non-negative.

Proof Let K be fixed. $D(K)$ is a linear subspace of dimension d in $\mathbb{R}^J$. Let $\vec{w} = \vec{w'} \in D(K)$ satisfy the conditions of the lemma. We define D_{Signs}, given by

$$D_{Signs} = \{\vec{u} \in D(K) \,|\, u_i = 0 \text{ for every } i \in T, u_i \geq 0 \text{ for every } i \text{ for which } w'_i > 0, u_i \leq 0 \text{ for every } i \text{ for which } w'_i < 0 \text{ and } \sum_{i=1}^{J} |u_i| = 1\} \tag{6.73}$$

In the presence of the conditions for non-negativity and non-positivity of co-ordinates, $\sum_{i=1}^{J} |u_i| = 1$ simply means $\sum_{i=1}^{J} (-1)^{sig_i} u_i = 1$ for some fixed $\vec{sig} \in \{0, 1\}^J$. By definition, D_{Signs} is a convex polytope.

Hence, by Lemma 3, the vertices of D_{Signs} are unique solutions of sets of equations of the form

$$K_{dir} + \left(\sum_{i=1}^{J} (-1)^{sig_i} u_i = 1 \right) + d - 1 \text{ equations of the form } u_i = 0 \tag{6.74}$$

i.e. a positive constant $\times$ some $\vec{v} \in V_0$. We assume that S, T are as in the lemma and that $w'_i \geq 0$ for every $i \in S$ and $w'_j = 0$ for every $j \in T$. Then for each $i = 1, \ldots J$, $w'_i \geq 0 \Leftrightarrow sig_i = 0$, so, if we define

$$\vec{w*} = \vec{w'} / \sum_{i=1}^{J} |w'_i| \tag{6.75}$$

then $\vec{w*}$ is a linear combination of the required form, so we have proved the lemma.

■

Lemma 123 *Let $Mxmn(K) = \vec{X}$, with $\vec{x}, stable$ etc. variables in the running of $MCA3$ with input K. Then at Step 4, $X_i = x_i$ for every $i \in stable$ and $V = \{\vec{v} \in V_0 \,|\, v_i = 0$ for all $i \in stable\}$ which equals the initial value of V if the input of the algorithm is $K + \{(x_i = X_i) | i \in stable\}$ instead of K.*

Proof We show that this is true at the m'th use of Step 4, by induction on m.

Base Case $\underline{m = 0}$ The $i \in stable$ are K-constant co-ordinates so $x_i = X_i$ and $V = V_0$.

Inductive Step Assume that the lemma holds at the previous use of Step 4. The $i \in minmobile$ are those i, of those not in $stable$, for which x_i is minimal. Hence $\vec{trydir}$ produces $maximin$-improvement from $\vec{x}$.

If possible, at Step 9, we use $\vec{U}$ such that those co-ordinates increase in value at the same rate, keeping them equal. Those $i \in minmobile$ for which $trydir_i = 0$ cannot be increased unless some of the members of $minmobile$ decrease from their $\vec{x}$ values. Hence $X_i = x_i$ because no direction giving $maximin$-improvement from $\vec{x}$ can change those co-ordinates' values so the set of them is denoted $newstable$ and they become members of $stable$. Also the directions $\vec{v} \in V$ that change them in value are removed. Thus, should the algorithm not terminate, the conditions of the lemma are true the next time Step 4 is reached and we have proved the Inductive Step and the lemma.

■

Lemma 124 *From any point in the running of MCA3, after a finite number of steps either stable increases or the algorithm terminates.*

Every time we use Step 4 we determine which non-*stable* co-ordinates have least value in $\vec{x}$, i.e. those in *minmobile* and whilst not decreasing any of them, we increase in value all that we can. Those that don't increase become members of *stable* and of those that do we check in Step 9 if there exists a "$\vec{U}$" i.e. if it is possible to increase them at the same rate.

Suppose for contradiction that the algorithm did not terminate with *stable* constant, and $newstable = \emptyset$. Let Min be one of the sets of largest cardinality which is a value of *minmobile* infinitely many times in the running of the algorithm.

Since there exist distinct values of y such that there are solutions of K such that every co-ordinate in Min takes the value y, there exists a direction in $D(K)$, $\vec{U}$, such that $U_i = 1$ for all $i \in Min$.

Hence if $minmobile = Min$, we find at Step 9 that K_{try} does have a real solution, say $\vec{U}$, and at Step 10 we move from $\vec{x}$ in the direction $\vec{U}$, so that at the new value of $\vec{x}$, it is still true that $x_i = x_j$ for all $i, j \in Min$. Thus the next value of *minmobile* must be a superset of Min.

However, there are only a finite number of possibilities for such a superset, so some $Min' \supset Min$ occurs an infinite number of times as the value of *minmobile* and has larger cardinality than Min, so we have found a contradiction. We have proved the lemma.

■

Proof of Algorithm 121 By Lemma 124, MCA 3 terminates at some $\vec{Y}$ since *stable* cannot increase more than J times. By Lemma 123, $\vec{Y}$ satisfies $K' = K + \{x_i = X_i \,|\, i \in stable\}$ and if there exist other solutions of K', $V \neq \emptyset$ by Lemma 122 so $\vec{Y} = \vec{X} = Mxmn(K)$. We have proved the algorithm.

■

6.6 *Maximin* Calculation Algorithm 4

Definition A co-ordinate i of a constraint set K is *gen-constant* (w.r.t. K) iff for some $c \in \mathbb{R}$ and all $\vec{x} \in G(K)$, $x_i = c$.

Algorithm 125 (*Maximin* **Calculation Algorithm 4)**

Input: *Some* $K \in \mathcal{C}_J^d$.

Output: $Mxmn(K)$.

1 Let $K* = K$, let $d* = d$ and define $allowed = \{1, 2, \ldots J\}$.

2 By row-reducing the constraints of $K*$, find all of the $q \in allowed$ which are gen-constant w.r.t. $K*$ and cast them out from *allowed*. If $d* = 0$, output the unique generalised solution of $K*$, $Mxmn(K)$, and stop.

3 Let *Possible* be the set of all vectors $\vec{x}$ which satisfy the following condition: $\vec{x}$ is the unique generalised solution of a constraint set of the form

$$K* + \{x_{i_1} = x_{i_2} = x_{i_3} = \ldots = x_{i_{d*+1}} \text{ s.t. all } i_p \in allowed\} \tag{6.76}$$

such that $x_{i_1} \leq x_r$ for all $r \in allowed$.

4 Then let $Best = \{\vec{x} \in Possible \text{ s.t. } min_{i \in allowed} x_i \text{ is maximal}\}$.

5 If *Best* is a singleton, output its member $= Mxmn(K)$ and stop.

6 Let $n_1, n_2, \ldots n_f$ be those co-ordinates in *allowed* which have the same value at every $\vec{x} \in Best$, say with values $V_1, \ldots V_f$ respectively. Let

$$K* = K* + \{x_{n_1} = V_1, x_{n_2} = V_2, \ldots x_{n_f} = V_f\} \tag{6.77}$$

delete $n_1, \ldots n_f$ from *allowed* and let $d* = dim(G(K*))$. Return to Step 2.

Proof We show that the variables above maintain certain properties using the following lemma.

Lemma 126 *Suppose that we run MCA 4 with input $K \in \mathcal{C}_J^d$ and that, at the end of Step 2, allowed $= \{i$ s.t. $i = 1, \dots J$ and i is not gen-constant w.r.t. $K*\}$. Then $Mxmn(K*)$ is a convex combination of the members of Best at the end of the next usage of Step 4.*

Proof Without loss of generality, let

$$Mxmn(K*) = (m_1, m_2, m_3, \dots m_f, c_1, c_2, \dots c_{con}) \tag{6.78}$$

where the gen-constant co-ordinates w.r.t. $K*$ are those listed with values $c_1, \dots c_{con}$ above and *allowed* $= \{1, \dots f\}$. We also let p be such that $m_1 = m_2 = m_3 = \dots = m_p < m_{p+1} \le m_{p+2} \le \dots \le m_f$. Let

$$K_*^{'} = K* \cup \{x_1 = m_1, x_2 = m_1, \dots x_p = m_1\} \tag{6.79}$$

so that $Mxmn(K*) = Mxmn(K_*^{'})$ since $Maximin$ satisfies Obstinacy (on constraint sets).

Consider the set

$$Pref = \{\vec{x} \in V(K_*^{'}) \text{ s.t. } x_i \ge m_1 \text{ for } i = 1, 2, \dots f\} \tag{6.80}$$

which is a convex polytope. By Lemma 3, each vertex of $Pref$ can be expressed as the unique solution of a constraint set of the form

$$K_*^{'} \cup \{x_{i_1} = m_1, x_{i_2} = m_1, \dots x_{i_g} = m_1\} \tag{6.81}$$

where $i_1, \dots i_g$ are such that $p + 1 \le i_1 < i_2 \dots < i_g \le f$. However,

$$K_*^{'} \equiv K \cup \{x_1 = x_2 = \dots = x_p\} \tag{6.82}$$

since if there exists $\vec{X} \in V(K*)$ s.t. $X_1 = \dots = X_p = \lambda \neq m_1$ there exists a direction $\vec{dir}$ parallel to $G(K*)$ in which all of those co-ordinates have positive change so $\vec{dir}$ produces $maximin$-improvement from $Mxmn(K*)$, which is a contradiction. Hence the constraint sets of the form (6.81) can be given equivalently by

$$K* \cup \{x_1 = x_2 = x_3 = \dots = x_p = x_{i_1} = \dots = x_{i_g}\} \tag{6.83}$$

Now $Mxmn(K*) \in P$ so to prove the lemma we need to see that each system (6.83) with a unique solution for which m_1 is the minimum value of the co-ordinates that are not gen-constant is in the final value of $Best$. This follows from carrying out Step 4, so $Mxmn(K*) \in P$ is a convex combination of those unique solutions as required and we have proved the lemma.

■

Proof of Algorithm 125 continued Let K be as in the input of MCA4. During the running of MCA4 at Step 2 it is always true that

$$Mxmn(K) = Mxmn(K*) \tag{6.84}$$

and *allowed* is the set of co-ordinates of $K*$ that are not gen-constant w.r.t. $K*$.

Then by Lemma 126, after Step 4, $Mxmn(K)$ is a convex combination of the members of $Best$, which all have the largest possible minimum of the non-gen-constants w.r.t. $K*$, m_1. Thus $Mxmn(K*) = Mxmn(K)$ at Step 6, since $Mxmn(K)$ remains a solution of $K*$. Now those co-ordinates with value m_1 at $Mxmn(K)$ are now gen-constant w.r.t. $K*$ but they weren't before, so every time we complete Step 2 some co-ordinates have been deleted from *allowed* and $d* = dim(K*)$ has decreased since our last usage of Step 2. This forces the algorithm to terminate and output $Mxmn(K)$.

■

6.7 Computational complexity

In this work we do not carry out a rigorous study of the computational complexity of the different algorithms mentioned in this chapter. However, below we make a few general remarks which can be the starting point of further research.

For the $Minimax$ Calculation Algorithm, let the input be $K \in \mathcal{C}_J^d$. There are $\frac{J^2+J}{2}$ EZ-constraints that can be added to K. Thus the number of constraint sets considered at Step 2 is bounded by $\begin{pmatrix} \frac{J^2+J}{2} \\ d \end{pmatrix}$.

Definition Let n and k be positive integers such that $k \leq n$. The *Stirling number of the second kind*, denoted by $S(n,k)$, is the number of equivalence relations on the set $\{1, 2, \ldots n\}$ which have exactly k equivalence classes, and it is given by

$$S(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \begin{pmatrix} k \\ i \end{pmatrix} (k-i)^n \tag{6.85}$$

If we modified the algorithm (as suggested half way down page 145) by using a list of all the possible equivalence relations on $\{0\} \cup \{$ co-ordinates $1, 2 \ldots J\}$ which can be expressed by the collections of added EZ-constraints, the number of such equivalence relations we would use is bounded by $S(J+1, d+1)$.

In terms of computational complexity, the $Maximin$ Calculation Algorithm Prototype is essentially similar to the $Minimax$ Calculation Algorithm.

The $Maximin$ Calculation Algorithm 1 seems to be a slight improvement on the Prototype. Instead of adding $\begin{pmatrix} \frac{J^2+J}{2} \\ d \end{pmatrix}$ sets of EZ-constraints, we use $\begin{pmatrix} \frac{J^2-J}{2} \\ d \end{pmatrix}$ sets of E-constraints. If, similarly to the above, we modified the algorithm so that the equivalence relations on $\{$ co-ordinates $1, 2, \ldots J\}$ were considered, the number of relations we would use is bounded by $S(J, d+1)$.

$Maximin$ Calculation Algorithm 2 traverses from an initial point, eventually reaching the $Maximin$ solution. In practice, it can sometimes be very fast, if only because we can "get lucky" and only have a small number of loops of the algorithm to traverse.

When experiments are carried out in using $MCA3$ with different inputs, our progress to the *Maximin* solution can be thought of as curiously fast.

Finally, $MCA4$ uses equivalence relations for which all but one of the equivalence classes are singletons. At the first usage of Step 3, the number of constraint sets considered is bounded by $\begin{pmatrix} J \\ J/2 \end{pmatrix}$. With each repetition of step 6, the cardinality of *allowed* is reduced. Hence, where $J = 2Q$, the total number of equivalence relations used in the algorithm is bounded by

$$Bound = \begin{pmatrix} 2Q \\ Q \end{pmatrix} + \begin{pmatrix} 2Q-1 \\ Q \end{pmatrix} + \begin{pmatrix} 2Q-2 \\ Q-1 \end{pmatrix} + \ldots + \begin{pmatrix} 2 \\ 1 \end{pmatrix} \tag{6.86}$$

and, since $\begin{pmatrix} 2Q-1 \\ Q \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2Q \\ Q \end{pmatrix}$,

$$\frac{2}{3} Bound \leq \sum_{i=1}^{Q} \begin{pmatrix} 2i \\ i \end{pmatrix} \tag{6.87}$$

Now it can be seen that $\begin{pmatrix} 2Q-2 \\ Q-1 \end{pmatrix} = \frac{Q}{4Q-2}\begin{pmatrix} 2Q \\ Q \end{pmatrix}$ so that the term of (6.87) given by $i = k$ is bounded by one half of the term given by $i = k+1$ and, consequently, $Bound \leq 3\begin{pmatrix} 2Q \\ Q \end{pmatrix}$. Hence the number of equivalence relations used at Step 3 during an entire execution of MCA 4 is bounded by $O(\begin{pmatrix} J \\ J/2 \end{pmatrix})$.

Chapter 7

Partly Linear inference processes

7.1 What is a PL inference process?

In [ParVen1], the authors outline desirable properties possessed by $Maximum\ Entropy$, before proving that ME is the only inference process satisfying all of them. However ME is not an easy inference process to calculate exactly, as the point minimising $\sum_{i=1}^{J} x_i \log(x_i)$ must be found, which gives us algebraic answers when the constraints use just rational coefficients. One advantage possessed by $Maximin$ is that, by the $Maximin$ Calculation Algorithm (Algorithm 101), it gives rational values to the inferred beliefs when the knowledge is expressed purely in terms of rational coefficients, so that it is "simpler" for us to write down exact answers.

By Chapter 5, $Maximin$ also satisfies Atomic Renaming, Equivalence, Obstinacy, Relativisation, Open-mindedness, Irrelevant Information, Language Invariance and Piecewise Linear Loaf Continuity. We might therefore hope to be able to express $Maximin$ as the unique inference process that satisfies all of the properties on a slightly different list from the list of those satisfied by ME.

The type of algorithm used to calculate $Maximin$ is generalised by the following definition:

Definition If $K \in CL$, $dim(V^L(K))$ denotes the number of dimensions of the smallest affine space in $\mathbb{R}^J$ which contains $V^L(K)$. We use the letter d in this chapter to denote $dim(V^L(K))$, as opposed to $dim(G(K))$ in Chapter 6.

Definition A partly linear (PL) inference process N^L over a language L is one for which there exist distinct $t_1, t_2, \ldots t_p$ s.t. for each $k = 1, \ldots p$,

$$t_k = \left\{ \vec{x} \in \mathbb{D}^J \text{ s.t. } \sum_{i=1}^{J} a_{ki} x_i = b_k \right\} \neq \mathbb{D}^J \tag{7.1}$$

for constants a_{ki}, b_k, and for every consistent knowledge base K such that $dim(V^L(K)) = d > 0$, there exist distinct $t_{i_1}, \ldots t_{i_d}$ with $1 \leq i_1, i_2, \ldots i_d \leq p$ such that

$$\{N^L(K)\} = V^L(K) \cap \bigcap_{j=1}^{d} t_{i_j} \tag{7.2}$$

(see [1] below). In such a case we say that $T = \{t_1, \ldots t_p\}$ is a *toolbox* which *allows* (us to calculate) N^L and the t_i's are *tools*. Tools may be denoted by constraints whose solution set they are, on the understanding that J is known and that, for example, $(x_1 = x_2)$ is the *same* as $(2x_2 = 2x_1)$ since both expressions are abbreviations for the tool

$$\left\{ \vec{x} \in \mathbb{D}^J \text{ s.t. } x_1 = x_2 \right\} \tag{7.3}$$

If, for $K \in CL$, $N^L(K)$ is characterised by an equation such as (7.2), we say that the tools $t_{i_1}, \ldots t_{i_d}$ *pick out* $N^L(K)$. We use these terms in Chapters 7 and 8 of this thesis.

Remark Linear functions can be thought of as having the advantage of being easier to calculate, and it can be argued that this makes PL inference processes more natural, in that if a rational agent is subconsciously making calculations, they may solve a series of linear equations before comparing the possible answers in some "simple" way, rather than choosing to minimise a non-linear function in a region of $\mathbb{D}^J$.

Definition For any overlying language L, we define the toolbox

$$T_0^L = \{(x_i = x_j) \text{ s.t. } 1 \leq i < j \leq J\} \tag{7.4}$$

[1] If $dim(V^L(K)) = 0$, $V^L(K)$ is a singleton, so $\{N^L(K)\} = V^L(K)$ must hold for *any* inference process N^L.

and let the toolbox T^L_{0+} be given by

$$T^L_{0+} = T^L_0 \cup \{(x_i = 0) \text{ s.t. } 1 \leq i \leq J\} \tag{7.5}$$

Theorem 127 *For all overlying languages L, $Minimax^L$ and $Maximin^L$ are partly linear. T^L_{0+} is a toolbox allowing $Minimax^L$ and T^L_0 allows $Maximin^L$.*

Proof We consider a specific $K \in CL$ such that $d = dim(V^L(K))$. By the *Minimax* Calculation Algorithm (Algorithm 94), $\{Mmx(K)\}$ is the unique generalised solution of a constraint set of the form

$$K \ + con_1, con_2, \ldots con_{dim(G^L(K))} \tag{7.6}$$

where the con_i are E-Z constraints. Now the set of p.b.f.'s satisfying any particular E-Z constraint is a tool in T^L_{0+}. If we fix $S \subseteq \{con_1, \ldots con_{dim(G^L(K))}\}$ to be of minimal cardinality such that

$$\{Mmx(K)\} = V^L(K + \{con_i \text{ s.t. } i \in S\}) \tag{7.7}$$

then S has d members and the corresponding set of tools pick out $Mmx(K)$. Thus $Minimax^L$ is partly linear and T^L_{0+} allows $Minimax^L$.

Similarly, by the *Maximin* Calculation Algorithm 1(Algorithm 108), we see that $Maximin^L$ is partly linear and T^L_0 allows $Maximin^L$.

■

Notation Since $\sum_{i=1}^J x_i = 1$ is true throughout $\mathbb{D}^J$, the tool $(\sum_{i=1}^J a_i x_i = c)$ can be written in the form

$$\left(\sum_{i=1}^{J} (a_i - c) x_i = 0 \right) \tag{7.8}$$

and we shall do this to ease notation. However, some tools do not have a unique expression of this form.

Hence in calculating a PL inference process using a given toolbox there is a finite list of extra constraints to choose from, and when enough are added to K the possible unique solutions include $N^L(K)$. In general one or more of the t_k's could be redundant i.e. the toolbox $T' = T \setminus t_k$ might also allow N^L. It is natural however to consider toolboxes which have no redundancies. A strong form of such minimality is given by the following definition.

Definition We say that a toolbox T is *uniquely minimal* for the PL inference process N^L iff T is a toolbox that allows N^L and for any toolbox T' allowing N^L, $T \subseteq T'$.

A PL inference process may fail to have a uniquely minimal toolbox T but any PL inference process has at most one uniquely minimal toolbox.

Remark The same toolbox can allow many useful inference processes. For example, for any language L, T^L_{0+} allows both $Maximin^L$ and $Minimax^L$, by Theorem 127. To calculate N^L, we need to know how to choose $N^L(K)$ from the possible unique solutions after adding the d constraints. Also a Language Invariant family of inference processes is often referred to as a single inference process and, though the toolboxes are finite for each overlying language, there might not necessarily be a natural way of naming them all simultaneously.

We now introduce the properties of Irrelevant Certainty and Homogeneity so that we can prove uniqueness results for $Maximin$.

Definition A language invariant family of inference processes N satisfies $Irrelevant\ Certainty$ iff for all $K \in CL$, if $K' = K + (Bel(p') = 0)$, where $p' \notin L$, $N(K)(\alpha) = N(K')(\alpha)$ for every $\alpha \in At^L$. We refer to this weakening of Irrelevant Information in Chapters 7 and 8.

Theorem 128 *Let N be a Language Invariant family of inference processes satisfying Atomic Renaming and Equivalence. Then N satisfies Irrelevant Certainty iff for all languages L (where $J = |At^L|$ as usual) and each $K \in CL$, if $N(K) = \vec{X}$ and K' is given by the constraints of K on $x_1, \ldots x_J$ together with*

$$x_{J+1} = x_{J+2} = \ldots = x_{2J} = 0, \ \ \textit{then} \ \ N(K') = (X_1, \ldots X_J, 0, 0, \ldots 0) \tag{7.9}$$

Proof (⇒): Suppose that N is a Language Invariant family of inference processes satisfying Atomic Renaming, Equivalence and Irrelevant Certainty. For some language L, we let $At^L = \{\alpha_1, \ldots \alpha_J\}$ and $x_i = Bel(\alpha_i)$ as usual. For some $K \in CL$, let $K' = K + \{x_{J+1} = 0, x_{J+2} = 0 \ldots x_{2J} = 0\}$. The $2J$ atoms β_i of $L' = L + p'$ (where p' is not in L) can be labelled

$$\beta_i = \alpha_i \wedge \neg p',\ \beta_{J+i} = \alpha_i \wedge p' \ \text{ for } \ i = 1, \ldots J \tag{7.10}$$

w.l.o.g., since N satisfies Atomic Renaming. Now K' can be equivalently expressed as K'' which is the result of including $\sum_{i=1}^{J} x_{J+i} = 0$ and the constraints of K with, for every $i = 1, \ldots J$, every occurrence of x_i replaced by $x_i + x_{J+i}$. However

$$x_i + x_{J+i} = Bel(\alpha \wedge \neg p') + Bel(\alpha \wedge p') = Bel(\alpha) \tag{7.11}$$

so K'' is equivalent to $K + Bel(p') = 0$. We let

$$N(K') = N(K'') = \vec{w} \in \mathbb{D}^{2J} \tag{7.12}$$

By Irrelevant Certainty, $N(K'')(\alpha) = N(K)(\alpha)$ for any $\alpha \in At^L$ so $w_i + w_{J+i} = X_i$ and, as $w_{J+i} = 0$, $w_i = X_i$ for every $i = 1, \ldots J$ as required.

(⇐): We can follow the steps of the above proof in reverse order. Suppose that N is an inference process satisfying the conditions of the hypotheses and the condition (7.9) is always satisfied, where $\alpha_1, \ldots \alpha_J$ enumerates the atoms of some language L.

If $K \in CL$ and $p' \notin L$, let $K' = K + \{Bel(p') = 0\}$. For all $i = 1, \ldots J$, we can label $\beta_{J+i} = \alpha_i \wedge p'$ and $\beta_i = \alpha_i \wedge \neg p'$ w.l.o.g., by Atomic Renaming. Then $\beta_1, \ldots \beta_{2J}$ enumerates the atoms of $L + p'$. We can rewrite K', up to equivalence, in the form

$$K' = K + \{Bel(p') = 0\} \tag{7.13}$$

so the fact that (7.9) holds shows that Irrelevant Certainty is satisfied. We have proved the theorem.

■

Theorem 129 *Suppose that N is a Language Invariant family of inference processes which satisfy Equivalence and Atomic Renaming. For all languages L, let N^L be given by*

$$N^L(K) = \text{ the unique } \vec{x} \text{ for which } \sum_{i=1}^{J} f(x_i) \text{ is minimal} \tag{7.14}$$

where the function $f : [0,1] \to \mathbb{R}$ is independent of J, where $J = |At^L|$.
Then N satisfies Irrelevant Certainty. Hence all Renyi Processes satisfy Irrelevant Certainty.

Also, for other reasons, CM_∞, $Minimax$, and $Maximin$ satisfy this property.

Proof By inspecting the definitions of the form (7.14) or the definition of CM_∞, we see that the condition (7.9) holds for these inference processes so Theorem 128 shows that Irrelevant Certainty is satisfied. Also for $Minimax$ and $Maximin$ we can see that (7.9) is always true, so we have proved the theorem.

■

Definition An inference process N^L is *Homogeneous* if whenever $\phi, \theta_{ji} \in SL$ and constants a_{ji}, b_j are fixed such that the knowledge bases

$$K_\lambda = \{Bel(\phi) = \lambda\} \cup \left\{ \sum_{i=1}^{s} a_{ji} Bel(\theta_{ji}|\phi) = b_j \text{ for } j = 1, \dots d \right\} \tag{7.15}$$

are consistent for all λ such that $0 \leq \lambda \leq 1$ (see [2] below), then

$$N^L(K_\lambda)(\theta|\phi) = N^L(K_1)(\theta|\phi) = N^L(K)(\theta) \tag{7.16}$$

where K denotes K_1, for all $\lambda \in [0,1]$ and all $\theta \in SL$. We refer to this property in Chapters 7 and 8.

Theorem 130 *If N is a family of Language Invariant inference processes satisfying Atomic Renaming, Obstinacy, Equivalence, Irrelevant Certainty, Relativisation and Piecewise Linear Loaf Continuity, then N is Homogeneous.*

Proof We assume that N is a family of inference processes satisfying the hypotheses of the theorem and that we have fixed a set of knowledge bases K_λ as in the

[2] We assume that $Bel(\theta|\phi) = b$ is shorthand for $Bel(\theta \wedge \phi) = bBel(\phi)$, so that K_0 is well-defined.

definition of Homogeneity which we write in the form

$$K_\lambda = \left\{ \sum_{i=1}^{q} x_i = \lambda, \sum_{i=1}^{J} x_i = 1 \right\} \cup \left\{ \sum_{i=1}^{q} a_{ji} x_i = \lambda b_j \text{ for } j = 1, \ldots d \right\} \tag{7.17}$$

with a specific overlying language L. We let $K = K_1$. As usual x_i denotes $Bel(\alpha_i)$ for each $i = 1, \ldots J$, where $At^L = \{\alpha_1, \ldots \alpha_J\}$. For N^L to satisfy Homogeneity we require that $N(K_\lambda)_i = \lambda N(K)_i$ for each $i = 1, \ldots q$.

Without loss of generality, we may assume that $J \geq 2q$. For if not consider K'_λ given by adding the extra variables $x_{J+1}, \ldots x_{2J}$ and the constraints

$$x_{J+1} = x_{J+2} = \ldots = x_{2J} = 0 \tag{7.18}$$

to K_λ. Then by Theorem 128 $N(K'_\lambda)_i = N(K_\lambda)_i$ for each i s.t. $1 \leq i \leq J$. Hence if Homogeneity fails for the K_λ it also fails for the K'_λ.

Lemma 131 *For each $i = 1, \ldots q$ and any $\lambda \in [0, 1]$,*

$$N(K_{\lambda/2})_i = \frac{1}{2} N(K_\lambda)_i \tag{7.19}$$

Proof For this proof we assume that $\lambda = \lambda_0$. For K_{λ_0}, $K_{\lambda_0/2}$ as above, define K_* by

$$K_* = K_{\lambda_0/2} \cup \left\{ \sum_{i=q+1}^{2q} x_i = \lambda_0/2 \right\} \cup \left\{ \sum_{i=q+1}^{2q} a_{ji} x_i = b_j \lambda_0/2 \text{ for } j = 1, \ldots d \right\} \tag{7.20}$$

that is, $K_{\lambda_0/2}$ together with a copy of those constraints on $x_{q+1}, \ldots x_{2q}$. For each i such that $1 \leq i \leq q$, $N(K_*)_i = N(K_*)_{q+i}$ by Atomic Renaming and $N(K_*)_i = N(K_{\lambda_0/2})_i$ by Relativisation. Define

$$K_+ = \left\{ \sum_{i=1}^{q} (x_i + x_{q+i}) = \lambda_0 \right\} \cup \left\{ \sum_{i=1}^{q} a_{ji}(x_i + x_{q+i}) = \lambda_0 b_j \text{ for } j = 1, \ldots d \right\} \tag{7.21}$$

Then by Atomic Renaming, $N(K_+)_i = N(K_+)_{q+i}$ for $i = 1, \ldots q$ so using Obstinacy,

$$N(K_+) = N(K_+ \cup \{x_i = x_{q+i} \text{ s.t. } i = 1, \ldots q\}) \tag{7.22}$$

but we see that

$$V(K_+ \cup \{x_i = x_{q+i} \text{ s.t. } i = 1, \ldots q\}) \subseteq V(K_*) \tag{7.23}$$

and $N(K_*)$ is a solution of K_+ so by Obstinacy $N(K_+) = N(K_*)$. The atoms of L are such that x_i denotes $Bel(\alpha_i)$ for each i, according to the chosen enumeration of the atoms α_i. By Atomic Renaming, if we change that enumeration, the inferred beliefs are permuted in the corresponding way.

Choose $p' \in L$ and assume w.l.o.g. that the atoms of $L\backslash p'$ are enumerated such that $\beta_1, \beta_2, \ldots \beta_q \in At^{L\backslash p'}$. We label $x_i = Bel(\beta_i \wedge p')$ and $x_{q+i} = Bel(\beta_i \wedge \neg p')$ for $i = 1, \ldots q$. Then we rewrite K_+:

$$K_+ = \left\{ \sum_{i=1}^{q} Bel(\beta_i) = \lambda_0 \right\} \cup \left\{ \sum_{i=1}^{q} a_{ji} Bel(\beta_i) = \lambda_0 b_j \ \text{ for } \ j = 1, \ldots d \right\} \tag{7.24}$$

Now, using the overlying language $L\backslash p'$, we let $y_i = Bel(\beta_i)$ for all $i = 1, \ldots \frac{J}{2}$ and we can now rewrite K_+ as

$$\left\{ \sum_{i=1}^{q} y_i = \lambda_0 \right\} \cup \left\{ \sum_{i=1}^{q} a_{ji} y_i = \lambda_0 b_j \ \text{ s.t. } \ 1 \le j \le d \right\} \tag{7.25}$$

so by Theorem 128 $N(K_+)(\beta_i) = N(K_{\lambda_0})(\alpha_i)$ for each i s.t. $i \le q$. Now

$$N(K_+)(\beta_i) = N(K_*)(\alpha_i \vee \alpha_{q+i}) = 2N(K_{\lambda_0/2})_i \tag{7.26}$$

for each $i = 1, \ldots q$ so $N(K_{\lambda_0/2})_i = 1/2 N(K_{\lambda_0})_i$ for those values of i and we have proved the lemma.

■

Proof of Theorem 130 continued

Now we let the loaf $\hat{K}$ (recall (3.43)) be given by

$$\hat{K} = \{ K_\lambda \,|\, 0 \le \lambda \le 1 \} \tag{7.27}$$

with slices K_λ and λ-range $[0, 1]$. By assumption, N is Piecewise Linear Loaf Continuous so for some $\delta = 2^{-k}$, where $k \in \mathbb{N}$, and some $\vec{X} \in V^L(K_0)$ and $\vec{u} \in \mathbb{R}^J$,

$$N(K_\lambda) = \vec{X} + \lambda \vec{u} \tag{7.28}$$

for all $\lambda \in [0, \delta]$. Since $x_1 + x_2 \ldots + x_q = 0$ is a constraint of K_0, $X_1 = \ldots = X_q = 0$. Hence for each $i = 1, \ldots q$, and all λ such that $\lambda \le \delta$, $N(K_\lambda)_i = \lambda u_i$.

Claim(w) For each $i = 1, \dots q$ $N(K_\lambda)_i = \lambda u_i$ for every λ s.t. $0 \leq \lambda \leq 2^{w-k}$.

Proof of claim We prove this by induction on $w = 0, \dots k$, using Lemma 131.

Base Case $\underline{w = 0}$ Since $\delta = 2^{-k}$, $N(K_\lambda)_i = \lambda u_i$ for any $\lambda \leq 2^{0-k}$ and for each $i = 1, \dots q$.

Inductive Step Assume that (Inductive Hypothesis) $N(K_\lambda)_i = \lambda u_i$ for all $i = 1, \dots q$ and $\lambda \in [0, 2^{w-k}]$. If

$$\lambda \in [0, 2^{w+1-k}],\ \lambda/2 \in [0, 2^{w-k}] \tag{7.29}$$

so $N(K_{\lambda/2})_i = u_i\lambda/2$ and by Lemma 131, $2N(K_{\lambda/2})_i = \lambda u_i = N(K_\lambda)_i$, completing the Inductive Step. The linear formula for $N(K_\lambda)_i$ holds for $\lambda \in [0, 2^{w+1-k}]$.

Hence for all $\lambda \in [0, 1]$ and every $i = 1, \dots q$, $N(K_\lambda)_i = \lambda u_i$, proving the claim.

Thus N is Homogenous and we have proved the theorem.

■

Corollary 132 *Minimax and Maximin satisfy Homogeneity.*

Proof Due to various results proved in Chapter 3 and Chapter 5, $Minimax$ and $Maximin$ satisfy the hypotheses of Theorem 130. Hence the corollary follows.

■

Theorem 133 *All of the Renyi Processes satisfy Homogeneity.*

Remark Since the Ren_r processes for which $r \neq 2$ do not satisfy Piecewise Linear Loaf Continuity, we cannot use Theorem 130 to deduce Homogeneity for these inference processes.

Proof We first consider the case of Ren_R, where R is a fixed real number and $R > 1$.

We fix a set of knowledge bases K_λ as in the definition of Homogeneity which we write, as in (7.17), in the form

$$K_\lambda = \left\{ \sum_{i=1}^{q} x_i = \lambda, \sum_{i=1}^{J} x_i = 1 \right\} \cup \left\{ \sum_{i=1}^{q} a_{ji} x_i = \lambda b_j \text{ for } j = 1, \dots d \right\} \tag{7.30}$$

with a specific overlying language L, and we define K to be K_1. As usual x_i denotes $Bel(\alpha_i)$ for each $i = 1, \ldots J$, where $At^L = \{\alpha_1, \ldots \alpha_J\}$. We need to prove that $Ren_R(K_\lambda)_i = \lambda Ren_R(K)_i$ for all $\lambda \in [0, 1]$ and each $i = 1, \ldots q$. The case $\lambda = 0$ is trivial.

Notation In general, let $(\vec{a}, \vec{b}, c, d, ...)$ be shorthand for the vector

$$(a_1, a_2, \ldots a_m, b_1, \ldots b_n, c, d, ...) \tag{7.31}$$

where $\vec{a} \in \mathbb{R}^m, \vec{b} \in \mathbb{R}^n$ etc.. We use this in Chapters 7 and 8.

By definition of the K_λ, if

$$V^L(K) = \{\vec{x} = (\vec{z}, 0, 0, 0, \ldots 0) \text{ s.t. } \vec{x} \in \mathbb{D}^J \text{ and } \vec{z} \in P\} \tag{7.32}$$

where P is a convex polytope which is a subset of $\mathbb{D}^q$, then for all K_λ,

$$V^L(K_\lambda) = \{\vec{x} = (\lambda\vec{z}, x_{q+1}, \ldots x_J) \text{ s.t. } \vec{x} \in \mathbb{D}^J \text{ and } \vec{z} \in P\} \tag{7.33}$$

In particular, $x_{q+1} + \ldots + x_J = 1 - \lambda$ is implied by the above condition and Ren_R will make these x_i all equal to $\frac{1-\lambda}{J-q}$ by Atomic Renaming. Let $Ren_R(K) = (\vec{Z}, 0, 0, 0 \ldots, 0)$, where $\vec{Z} \in P$.

Suppose for contradiction that, for some $\lambda_0 \in (0, 1]$ and some $\vec{Z'} \in P$ not equal to $\vec{Z}$,

$$Ren_R(K_{\lambda_0}) = \left(\lambda\vec{Z'}, \frac{1-\lambda}{J-q}, \ldots \frac{1-\lambda}{J-q}\right) \tag{7.34}$$

Since $Ren_R(K) = (\vec{Z}, 0, \ldots 0)$ and $(\vec{Z'}, 0 \ldots, 0) \in V^L(K)$,

$$\sum_{i=1}^{q} Z_i^R < \sum_{i=1}^{q} {Z'}_i^R \tag{7.35}$$

However, multiplying (7.35) by λ_0^R and adding $(J-q)\left(\frac{1-\lambda_0}{J-q}\right)^R$ shows that

$$\left(\lambda_0\vec{Z}, \frac{1-\lambda}{J-q}, \ldots \frac{1-\lambda}{J-q}\right) \tag{7.36}$$

has a smaller value of $\sum_{i=1}^{J} x_i^R$ than $Ren_R(K_{\lambda_0})$, which is a contradiction. Hence the Ren_r for which $r > 1$ satisfy Homogeneity. By a very similar argument we see that the Ren_r for which $0 < r < 1$ also satisfy this property.

Thus for every r s.t. $0 < r < 1$ and all $\lambda \in [0, 1]$,

$$Ren_r(K_\lambda)(\theta|\phi) = Ren_r(K_1)(\theta|\phi) \tag{7.37}$$

for all $\theta \in SL$, where $\phi = \bigwedge_{i=1}^{q} \alpha_i$ as in (7.17). Taking limits as $r \nearrow 1$ gives

$$ME(K_\lambda)(\theta|\phi) = ME(K_1)(\theta|\phi) \tag{7.38}$$

by [Moh]. Hence *Maximum Entropy* satisfies Homogeneity and we have proved the theorem.

■

Theorem 134 *CM_∞ satisfies Homogeneity.*

Proof We fix a set of knowledge bases K_λ as in the proof of Theorem 133. If we now take limits of (7.37) as $r \searrow 0$ we obtain

$$CM_\infty(K_\lambda)(\theta|\phi) = CM_\infty(K_1)(\theta|\phi) \tag{7.39}$$

by Theorem 32, so CM_∞ also satisfies Homogeneity and we have proved the theorem.

■

7.2 Showing that some PL inference processes have uniquely minimal toolboxes

For some PL inference processes N^L there may not exist a uniquely minimal toolbox. However in this section we show that uniquely minimal toolboxes exist for $Minimax^L$ and $Maximin^L$, for all overlying languages L. When we calculate an inference process N^L that satisfies Obstinacy and Atomic Renaming it is useful to observe symmetries in the knowledge base K and if, for example, interchanging x_1 and x_2 in the constraints of K does not change $V^L(K)$, then $x_1 = x_2$ is true at $N^L(K)$. Hence it should not surprise us that the tools $(x_i = x_j)$ can be found in toolboxes which allow N^L but, as the following theorem shows, it is impossible to avoid using them!

Theorem 135 *If N^L, allowed by a toolbox T, is a PL inference process satisfying Obstinacy, Equivalence and Atomic Renaming, $(x_i = x_j) \in T$ for all i, j s.t. $1 \leq i < j \leq J$.*

Proof To prove this we consider a selection of knowledge bases at which we can find the value of N^L just by assuming the properties of N^L given by the hypotheses of the theorem. We let Λ be given by

$$\Lambda = \left\{ \vec{\lambda} \in \mathbb{R}^{J-2} \text{ s.t. } \vec{\lambda} \text{ is non-negative and } \lambda = \sum_{i=1}^{J-2} \lambda_i < 1 \right\} \tag{7.40}$$

and define, for all $\vec{\lambda} \in \Lambda$,

$$K_{\vec{\lambda}} = \left\{ x_1 = \lambda_1, x_2 = \lambda_2, \ldots x_{J-2} = \lambda_{J-2}, \sum_{i=1}^{J} x_i = 1 \right\} \tag{7.41}$$

which is a function of the variable $\vec{\lambda}$.

Notation Let $(\vec{a} \cdot \vec{x} = 0)$ denote the tool $(\sum_{i=1}^{J} a_i x_i = 0)$, where $\vec{a} \in \mathbb{R}^J$. Then

$$V^L(K_{\vec{\lambda}}) = \left\{ (\vec{\lambda}, c - \tau, c + \tau), \text{ where } c = \frac{1}{2}(1 - \lambda) \text{ and } -c \leq \tau \leq c \right\} \tag{7.42}$$

and by symmetry between x_{J-1} and x_J, $N^L(K_{\vec{\lambda}}) = (\vec{\lambda}, c, c)$. Since N^L is partly linear, one tool from T, say $(\vec{a}^{\vec{\lambda}} \cdot \vec{x} = 0)$, dependant on $\vec{\lambda}$, is added to $K_{\vec{\lambda}}$ to pick out $N^L(K_{\vec{\lambda}})$.

Now

$$\lambda_1 a_1^{\vec{\lambda}} + \lambda_2 a_2^{\vec{\lambda}} + \ldots + \lambda_{J-2} a_{J-2}^{\vec{\lambda}} + \frac{1}{2}(1 - \lambda_1 - \ldots - \lambda_{J-2})(a_{J-1}^{\vec{\lambda}} + a_J^{\vec{\lambda}}) = 0 \qquad (7.43)$$

Rearranging (7.43) gives

$$\lambda_1(2a_1^{\vec{\lambda}} - a_{J-1}^{\vec{\lambda}} - a_J^{\vec{\lambda}}) + \lambda_2(2a_2^{\vec{\lambda}} - a_{J-1}^{\vec{\lambda}} - a_J^{\vec{\lambda}}) + \ldots + \lambda_{J-2}(2a_{J-2}^{\vec{\lambda}} - a_{J-1}^{\vec{\lambda}} - a_J^{\vec{\lambda}}) + a_{J-1}^{\vec{\lambda}} + a_J^{\vec{\lambda}} = 0 \qquad (7.44)$$

For every $\vec{\lambda} \in \Lambda$, some $\vec{a}^{\vec{\lambda}}$ for which $(\vec{a}^{\vec{\lambda}} \cdot \vec{x} = 0) \in T$ satisfies (7.44) because T allows N^L. Suppose for contradiction that the set of $\vec{\lambda}$ values covered by each tool in T (in the sense of (7.44)) can be contained in the intersection of a $(J-3)$-dimensional affine set and Λ. Then Λ is covered by a finite union of $(J-3)$-dimensional affine sets, so we have a contradiction. Hence for some $\vec{a} = \vec{A}$, $(\vec{a} \cdot \vec{x} = 0) \in T$ must satisfy (7.44) for a $(J-2)$-dimensional affine space of values of $\vec{\lambda} \in \mathbb{R}^{J-2}$ and hence for all $\vec{\lambda} \in \mathbb{R}^{J-2}$. By inspecting the cases of $\vec{\lambda} = (1, 0, 0, \ldots 0), \vec{\lambda} = (0, 1, 0, \ldots 0)$ etc. we see that

$$A_{J-1} + A_J = A_1 = A_2 = A_3 = \ldots = A_{J-2} = 0 \qquad (7.45)$$

so the tool used for these knowledge bases is $(x_{J-1} = x_J) \in T$. Similarly, by symmetry, $(x_i = x_j)$ must be in the toolbox for every i, j s.t. $1 \le i < j \le J$ and we have proved the theorem.

■

Corollary 136 *For any overlying language L, T_0^L is a uniquely minimal toolbox for $Maximin^L$.*

Proof By Theorem 127, T_0^L allows $Maximin^L$. Since $Maximin^L$ satisfies Obstinacy, Equivalence and Atomic Renaming (Chapter 5) every $(x_i = x_j)$ must be in all toolboxes that allow it, by Theorem 135. We have proved the corollary.

■

Theorem 137 *If L is a language such that $|L| > 1$, T_{0+}^L is a uniquely minimal toolbox for $Minimax^L$.*

Proof Let T be a fixed toolbox that allows $Minimax^L$, for some language L s.t. $|L| > 1$. Since $Minimax^L$ satisfies Obstinacy, Equivalence and Atomic Renaming (Chapter 3) all tools of the form $(x_i = x_j)$ must be in T by Theorem 135. Now we let M denote the set of non-negative $\vec{\mu} \in \mathbb{R}^{J-1}$ s.t. $\mu_1 > \mu_2 > 0$ and $\sum_{i=1}^{J-1} \mu_i = 1$. For all $\vec{\mu} \in M$, we define

$$K_{\vec{\mu}} = \left\{ x_1 = x_J + \mu_1, x_3 = \mu_3, x_4 = \mu_4, \ldots x_{J-1} = \mu_{J-1}, \sum_{i=1}^{J} x_i = 1 \right\} \tag{7.46}$$

Then

$$V^L(K_{\vec{\mu}}) = \{(\tau + \mu_1, \mu_2 - 2\tau, \mu_3, \ldots \mu_{J-1}, \tau) \text{ s.t. } 0 \leq \tau \leq \mu_2/2\} \tag{7.47}$$

$$= \{(\vec{\mu}, 0) + (\tau, -2\tau, 0, 0, \ldots 0, \tau) \text{ s.t. } 0 \leq \tau \leq \mu_2/2\} \tag{7.48}$$

using the notation from the proof of Theorem 135. We can calculate $Minimax$ by ignoring the $K_{\vec{\mu}}$-constant co-ordinates (Corollary 43) and the maximum value of the others is x_1, which is minimal when $\tau = 0$ so

$$Mmx(K_{\vec{\mu}}) = (\vec{\mu}, 0) \tag{7.49}$$

T allows Mmx^L so for all $\vec{\mu} \in M$, there exists $\vec{a}^{\vec{\mu}}$ for which $(\vec{a} \cdot \vec{x} = 0) \in T$ and

$$a_1^{\vec{\mu}} \mu_1 + a_2^{\vec{\mu}} \mu_2 + \ldots + a_{J-1}^{\vec{\mu}} \mu_{J-1} = 0 \tag{7.50}$$

We follow a similar argument to the proof of Theorem 135. Now M is $(J-2)$-dimensional and cannot be covered by a finite union of $(J-3)$-dimensional affine spaces. Thus, since T is finite, for some $\vec{a} = \vec{A}$, the fixed tool $(\vec{A} \cdot \vec{x} = 0)$ must satisfy (7.50) for a $(J-2)$-dimensional set of the $\vec{\mu}$ so that it picks out $Mmx^L(K_{\vec{\mu}})$. Hence it satisfies (7.50) for all $\vec{\mu} \in \mathbb{R}^{J-1}$ s.t. $\sum_{i=1}^{J-1} \mu_i = 1$. By inspecting the cases of $\vec{\mu} = (1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0) \ldots$ and $(0, 0, \ldots, 1)$ respectively,

$$A_1 = A_2 = \ldots = A_{J-1} = 0 \tag{7.51}$$

Hence the tool that picks out Mmx^L for these knowledge bases is $(x_J = 0)$. By symmetry, $(x_i = 0) \in T$ for every i such that $1 \leq i \leq J$ when T allows $Minimax^L$. By Theorem 127, T_{0+}^L allows $Minimax^L$ so we have proved the theorem.

∎

7.3 Some uniqueness results for $Maximin$

Since every Obstinate PL inference process satisfying Atomic Renaming must have all possible tools $(x_i = x_j)$ in its toolbox it is natural for us to define the simplest class of Partly Linear inference processes as follows:

Definition Given an overlying language L, a PL_0 inference process is one allowed by the toolbox $T_0^L = \{(x_i = x_j) \text{ s.t. } \ 1 \leq i < j \leq J\}$.

Theorem 138 *$Maximin$ is the only Language Invariant family of PL_0 inference processes satisfying Obstinacy, Equivalence, Atomic Renaming, Piecewise Linear Loaf Continuity, Irrelevant Certainty and Homogeneity.*

Proof By the results of Chapter 5 and Theorem 130, $Maximin$ satisfies the properties listed above. Now it remains for us to prove that $Maximin$ is uniquely characterised by those properties.

We fix N to be a Language Invariant family of PL_0 inference processes satisfying the above properties. Suppose for contradiction that for some overlying language L and $K' \in CL$, $N(K') = \vec{a} \neq \vec{b} = Mxmn(K')$ and let

$$V = \{\vec{x} = \vec{a} + \tau(\vec{b} - \vec{a}) \,|\, \tau \in \mathbb{R} \text{ and } \vec{x} \text{ is non-negative } \} \tag{7.52}$$

We now consider a knowledge base $K \in CL$ for which $V = V^L(K)$. We know that we can do this by Lemma 4. Then since N and $Maximin$ satisfy Obstinacy and $K' + K \equiv K$, $N(K + K') = N(K) = N(K')$ and $Mxmn(K) = Mxmn(K')$ so $N(K) \neq Mxmn(K)$.

Suppose that, for some language L, there exists $K \in CL$ s.t. $N(K) \neq Mxmn(K)$ and $V^L(K)$ is 1-dimensional, i.e. it is a line segment. The following two claims show that we can, w.l.o.g., assume certain things about K in order to prove Theorem 138.

Claim Without loss of generality, we can assume that $I^L(K) \neq \emptyset$.

Proof of claim Let $L' = L + p'$ for some $p' \notin L$ and let $K*$ be the constraints of K together with the extra co-ordinates $J+1, \ldots 2J$ and constraints

$$x_{J+1} = x_{J+2} = \ldots = x_{2J} = 0 \tag{7.53}$$

Now $Maximin^{L'}$ and $N^{L'}$ satisfy Irrelevant Certainty so invoking Theorem 128, we see that

$$Mxmn^{L'}(K*)_i = Mxmn^L(K)_i \text{ and } N^{L'}(K*)_i = N^L(K)_i \tag{7.54}$$

for each $i = 1, \dots J$ so $Mxmn(K') \neq N(K')$. Also $|I^{L'}(K*)| \geq J$ and $V^{L'}(K*) = \{(\vec{x}, 0, 0, \dots 0) | \vec{x} \in V^L(K)\}$, which is 1-dimensional.

■

Claim If K has a constant co-ordinate, say J w.l.o.g., with constant value $1-c$ then $\{N(K)\} = V^L(K) \cap (x_i = x_j)$ for some $i \neq J, j \neq J$.

Proof of claim Let $V^L(K) = \{$ non-negative $(c(\vec{v} + \tau\vec{w}), 1-c)$ s.t. $\tau \in \mathbb{R}\}$ where $\vec{w}$ is some fixed direction vector and $N(K) = (c\vec{v}, 1-c)$. We assume w.l.o.g. that K includes the constraint $\sum_{i=1}^{J-1} x_i = c$ and that none of the other constraints of K refer to x_J. We assume w.l.o.g. (by Atomic Renaming) that

$$K = \left\{ \sum_{i=1}^{J-1} x_i = c, \sum_{i=1}^{J} x_i = 1 \right\} \cup \left\{ \sum_{i=1}^{J-1} a_{ji} x_i = cb_j \text{ for } j = 1, \dots d \right\} \tag{7.55}$$

where the a_{ji}'s and b_j's are real constants. Define, for all $\mu \in [0, 1]$,

$$K_\mu = \left\{ \sum_{i=1}^{J-1} x_i = \mu, \sum_{i=1}^{J} x_i = 1 \right\} \cup \left\{ \sum_{i=1}^{J-1} a_{ji} x_i = \mu b_j \text{ for } j = 1. \dots d \right\} \tag{7.56}$$

Now by inspecting the constraints we see that

$$V^L(K_\mu) = \{\text{non-negative} \, (\mu(\vec{v} + \tau\vec{w}), 1-\mu)\} \tag{7.57}$$

so that $K_c \equiv K$ and by the Homogeneity of N,

$$N(K_\mu) = (\mu\vec{v}, 1-\mu) \tag{7.58}$$

for all μ such that $0 \leq \mu \leq 1$. If the tool used to pick out $N(K_\mu)$ is $(x_i = x_J)$ for some i s.t. $1 \leq i < J$ then

$$\mu v_i = (1-\mu) \tag{7.59}$$

Hence such a tool can only pick out $N(K_\mu)$ for one value of μ. There are infinitely many values of μ and each must have a tool in the finite toolbox T_0^L to pick out $N(K_\mu)$. Hence some tool $(x_i = x_j)$ for which neither i nor j equal J must work for some value of μ, say $\mu = \mu'$. However then $\mu' v_i = \mu' v_j$ so that condition must hold for all μ. Thus we can always use a tool of the form $(x_i = x_j)$ s.t. neither i nor j are K-constant, to pick out $N^L(K)$. Hence we have proved the claim.

■

Proof of Theorem 138 continued For the rest of this proof, we assume w.l.o.g. that K and L are fixed such that $V^L(K)$ is 1-dimensional, $I^L(K) \neq 0$, $N(K) \neq Mxmn(K)$ and a tool of the form $(x_i = x_j)$ picks out $N(K)$, where i and j are not K-constant co-ordinates.

Using Atomic Renaming, we can write w.l.o.g.,

$$V^L(K) = \{\vec{x} = (p_1 + q_1\tau,\ p_2 + q_2\tau, \ldots p_{l-1} + q_{l-1}\tau,\ \tau, t_1, t_2, \ldots t_r, 0)\}$$
$$\cap\{\vec{x} \in \mathbb{R}^J \text{ s.t. } x_i \geq 0 \text{ for every } i = 1, 2, \ldots J\} \tag{7.60}$$

where t_i, p_i, q_i are constants and the $q_i's$ are non-zero. Now for all real, non-negative λ, define

$$K_\lambda = \Big\{x_1 = p_1 + (q_1 - 1)\frac{x_J}{l} + q_1 x_l,\ x_2 = p_2 + (q_2 - 1)\frac{x_J}{l} + q_2 x_l, \ldots$$
$$x_{l-1} = p_{l-1} + (q_{l-1} - 1)\frac{x_J}{l} + q_{l-1} x_l,\ x_{l+1} = t_1,\ x_{l+2} = t_2, \ldots x_{l+r} = t_r, x_J = \lambda\Big\} \tag{7.61}$$

Let h be maximal such that K_h is consistent, then $\{K_\lambda | \, 0 \leq \lambda \leq h\}$ is a loaf and $K_0 \equiv K$. To write a solution of K_λ in a format similar to (7.61), let the parameter τ be given by $\tau = x_l + \frac{\lambda}{l}$. Then

$$V_\lambda = V^L(K_\lambda) = \Big\{\vec{x} \in \mathbb{R}^J \,|\, x_i \geq 0 \text{ for } i = 1, \ldots J \text{ and}$$
$$\vec{x} = \left(p_1 - \frac{\lambda}{l} + q_1\tau, p_2 - \frac{\lambda}{l} + q_2\tau, \ldots p_{l-1} - \frac{\lambda}{l} + q_{l-1}\tau, \tau - \frac{\lambda}{l}, t_1, \ldots t_r, \lambda\right)\Big\} \tag{7.62}$$

Each solution of K_λ, $\vec{x}$ is uniquely specified by the values of λ and τ as in (7.62). At $\vec{a} = N(K_\lambda)$, as at $\vec{b} = Mxmn(K_\lambda)$ two of the non-constant co-ordinates $1, \ldots l$ are equal in value which are not equal throughout V_λ as long as V_λ is infinite; this is because of the claim above, and the fact that N and $Mxmn$ are both PL_0. Let

$$G_\lambda = G^L(K_\lambda) \tag{7.63}$$

If we fix $i = i'$, $j = j'$ such that $1 \leq i' < j' \leq l$ either $x_{i'} = x_{j'}$ for every $\vec{x} \in G_\lambda$ (so for all λ, τ), $x_{i'} \neq x_{j'}$ in every case, or $x_{i'} = x_{j'}$ iff

$$\tau = m_{i'j'} = \frac{p_{i'} - p_{j'}}{q_{j'} - q_{i'}} \tag{7.64}$$

(where for x_l, define $p_l = 0, q_l = 1$) for every λ. Now given λ the range of values of τ for which $\vec{x}$ is non-negative is governed by the non-constant co-ordinates. As λ increases the lower bounds given by those i for which $q_i > 0$ increase and the upper bounds given by the cases $q_i < 0$ decrease until when $\lambda = h$ there exists a unique solution of K_h for which $\tau = X$, say.

By Piecewise Linear Loaf Continuity of N and of $Maximin$ the choices of m_{ij} , say respectively $N(\lambda)$ and $Mxmn(\lambda)$, must be continuous functions of λ. As there are finitely many m_{ij}'s, $N(\lambda)$ and $Mxmn(\lambda)$ are constant functions. As $\lambda \to h$,

$$N(\lambda) \to X \text{ and } Mxmn(\lambda) \to X \tag{7.65}$$

by the sandwich rule for limits but this is not possible unless $N = Maximin$ at K_0, contradicting the initial assumption. Hence $N = Maximin$ and we have proved the theorem.

■

Remark Surprisingly, the above proof, in effect, proves a uniqueness theorem for the listed properties of $Maximin$ which does not refer to the definition of $Maximin$ at all. Also, as we proved in Chapter 5, the inference process uniquely characterised by the theorem, i.e. $Maximin$, turns out to satisfy Irrelevant Information and Open-mindedness as well.

Corollary 139 *$Maximin$ is the only Language Invariant family of PL_0 inference processes satisfying Obstinacy, Equivalence, Atomic Renaming, Piecewise Linear Loaf Continuity, Relativisation, Open-mindedness and Irrelevant Information.*

Proof By the results of Chapter 5, we know that $Maximin$ satisfies each of the properties listed in the corollary. If some other family of Language Invariant inference processes exists, say N, satisfying these properties, then by Theorem 130 N satisfies Homogeneity. Also Irrelevant Information implies Irrelevant Certainty so N gives us a contradiction, in light of Theorem 138. Hence we have proved the corollary.

■

Remark *Maximin* also has another attractive feature. When calculating a general PL inference process N^L there might not be a simple algorithm which, when we input $K \in CL$, determines the tools that pick out $N^L(K)$, but for *Maximin* we need only compare numbers in size. Thus *Maximin* can justifiably claim to be the "best simple PL inference process there is".

Definition A PL_1 inference process is a partly linear inference process allowed by a toolbox containing only tools of the form $(a_1x_1 + a_2x_2 + \ldots + a_Jx_J = 0)$ for which $\sum_{i=1}^{J} a_i = 0$.

The following theorem provides a slight strengthening of Theorem 138.

Theorem 140 *Maximin is the only Language Invariant family of PL_1 inference processes satisfying Obstinacy, Equivalence, Atomic Renaming, Piecewise Linear Loaf Continuity, Irrelevant Certainty and Homogeneity.*

Proof This can be proved similarly to Theorem 138. Any tool used for a PL_1 inference process N is applied without involving a constant co-ordinate for essentially the same reason. Also the tools, like $(x_i = x_j)$, can be seen to contain a point $\vec{x}$ of the form (7.62) or not, independently of λ. Then Piecewise Linear Loaf Continuity guarantees that N and $Mxmn$ keep the same value of τ as λ varies, leading to the same contradiction as before if $N(K) \neq Mxmn(K)$.

■

7.4 The properties not satisfied by $Maximin$

We have found (see Chapter 5) that $Maximin$ satisfies all nine of the Par-Ven Properties apart from Independence and Continuity. The following theorems confirm that, among PL inference processes, $Maximin$ is the closest it can be to ME in terms of that list of desiderata.

Remark We should not find the following theorem surprising since for an inference process N^L to give the right answers to satisfy Independence requires it to show non-linear behaviour.

Theorem 141 *If $L = \{p_1, p_2, p_3\}$, no PL inference process N^L satisfies Independence.*

Proof Suppose that N^L is a PL inference process and satisfies Independence. Then for all knowledge bases of the form

$$K_{1,b,c} = \{Bel(p_1) = 1, Bel(p_2|p_1) = b, Bel(p_3|p_1) = c\} \tag{7.66}$$

for which $0 \leq b \leq c \leq 1/2$,

$$N^L(K_{1,b,c})(p_1 \wedge p_2 \wedge p_3) = bc \tag{7.67}$$

The standard ordering on the atoms of L leads to the expression

$$K_{1,b,c} = \{x_5 = x_6 = x_7 = x_8 = 0,\ x_1 + x_2 + x_3 + x_4 = 1,\ x_1 + x_2 = b,\ x_1 + x_3 = c\} \tag{7.68}$$

and, when we let $V(b, c)$ denote $V^L(K_{1,b,c})$,

$$V(b, c) = \{(\tau, b - \tau, c - \tau, 1 + \tau - b - c, 0, 0, 0, 0) \text{ s.t. } 0 \leq \tau \leq b\} \tag{7.69}$$

Let T be a toolbox which allows N^L and suppose that $(\vec{a} \cdot \vec{x} = 0)$ is the tool in T which picks out $N^L(K_{1,b,c})$, denoted by $Ind(b, c)$, which is the solution given by $\tau = bc$. Then

$$a_1 bc + a_2 b(1 - c) + a_3 c(1 - b) + a_4 (1 - c)(1 - b) = 0 \tag{7.70}$$

and a_i are not all zero for $i = 1, 2, 3, 4$, since if they are $(\vec{a} \cdot \vec{x} = 0)$ is not a tool.

Given $\vec{a}$, the set of values of b, c which satisfy satisfy (7.70) is a quadratic curve.

For every tool t in the toolbox allowing N^L, the set of values (b, c) for which t picks out $N^L(K_{1bc}) = Ind(b, c)$ is a quadratic curve. However the toolbox is finite, and the set of values (b, c) for which $0 \leq b \leq c \leq 1/2$ cannot be covered by a union of finitely many quadratic curves, so we have deduced a contradiction and proved the theorem.

■

Theorem 142 *If N^L is a PL inference process satisfying Equivalence and $|L| \geq 2$, N^L is not continuous.*

Remark In this thesis we only consider inference processes that satisfy Equivalence and Continuity makes no sense without it since the Blaschke distance $\Delta(K, K') = 0$ when $K \equiv K'$.

Proof Suppose for contradiction that N^L is a PL inference process satisfying Equivalence and Continuity and that $|L| \geq 2$. We introduce some notation just for this proof.

Notation Let Tri be the triangle with vertices

$$(1, 0, 0, 0, 0, \ldots 0), (0, 1, 0, 0, \ldots 0), (0, 0, 1, 0, 0, \ldots 0) \tag{7.71}$$

If $\vec{y}$ and $\vec{z}$ are distinct points on the boundary of Tri with vertices which do not lie on the same edge of Tri, we can, by Lemma 4 write a knowledge base which we denote by $K_{\vec{y},\vec{z}}$, determined up to Equivalence, for which $V^L(K_{\vec{y},\vec{z}})$ is the line segment connecting $\vec{y}$ and $\vec{z}$.

For any such $K_{\vec{y},\vec{z}}$, every co-ordinate apart from the first 3 is constant with value zero. The tools that pick out $N^L(K_{\vec{y},\vec{z}})$, of the form $(\vec{a} \cdot \vec{x} = 0)$, are identified in this proof purely by a_1, a_2, a_3, without loss of generality. From now on the p.b. functions are given by $\vec{x} \in \mathbb{D}^3$, so the co-ordinates $4, \ldots J$ are ignored since $x_4 = x_5 = \ldots = x_J = 0$ for every p.b. function $\vec{x}$ we consider until the end of this proof.

Definition If $\vec{X} \in \mathbb{D}^3$ is strictly positive and $\theta \in \mathbb{R}$, let

$$Line(\vec{X}, \theta) = \{\vec{x} = \vec{X} + \lambda(cos\theta, sin\theta, (-cos\theta - sin\theta)) \text{ s.t. } \vec{x} \in \mathbb{D}^3\} \tag{7.72}$$

θ is the angle in the x_1-x_2 plane of the direction parallel to the line segment $Line(\vec{X}, \theta)$ with the rate of change of x_3 chosen to be such that $x_1 + x_2 + x_3$ is constant, so that the end points of $Line(\vec{X}, \theta)$ are on the boundary of $\mathbb{D}^3$. Thus every line segment containing $\vec{X}$ which is the intersection of $\mathbb{D}^3$ and an infinite line is of the form $Line(\vec{X}, \theta)$.

Lemma 143 *If $\vec{X}$ is constant, $Line(\vec{X}, \theta)$ is a continuous function of θ in the Blaschke topology.*

Proof Let $\theta \to \theta_0 \in \mathbb{R}$. Now *sin*, *cos* are continuous and by inspecting (7.72), we see that for all θ, $\sum_{i=1}^{3} x_i = 1$ for every $\vec{x} \in Line(\vec{X}, \theta)$. Hence the set of values of λ for which the $\vec{x}$ in (7.72) are non-negative is the interval

$$[\lambda_{lower}(\theta), \lambda_{upper}(\theta)] \tag{7.73}$$

where λ_{lower} and λ_{upper} are continuous functions of θ bounded away from zero. As $\theta \to \theta_0$ the corresponding end points of $Line(\vec{X}, \theta)$ tend to those of $Line(\vec{X}, \theta_0)$. By Lemma 11, we complete the proof of this lemma.

■

Proof of Theorem 142 continued

Case 1 For every $K = K_{\vec{y},\vec{z}}$, $N^L(K)_i = 0$ for some $i = 1, 2$ or 3.

In this case, for every $K_{\vec{y},\vec{z}}$, $N^L(K_{\vec{y},\vec{z}}) = \vec{y}$ or $\vec{z}$. Consider those $K_{\vec{y},\vec{z}}$ for which $\vec{X} = (1/3, 1/3, 1/3)$ is a solution. The solution sets of these $K_{\vec{y},\vec{z}}$ are of the form $Line(\vec{X}, \theta)$, with θ taking all real values. The inference process is choosing an end point from each solution set, i.e. choosing $\lambda = \lambda_{lower}$ or $\lambda = \lambda_{upper}$ as in (7.72). If, when $\theta = 0$, the upper limit of λ, $\lambda_{upper}(0) = 1/3$, is chosen, then

$$N^L\left(x_2 = \frac{1}{3},\, x_1 + x_3 = \frac{2}{3}\right) = \left(\frac{2}{3}, \frac{1}{3}, 0\right) \tag{7.74}$$

However, when $\theta = \pi$, $\lambda = \lambda_{lower}(0) = -1/3$ must be chosen to get the same answer from the same knowledge base. We now see that this contradicts the continuity of

N^L. Indeed λ_{upper} and λ_{lower} vary continuously and are bounded away from zero as θ varies from 0 to π and λ_{upper} is positive, λ_{lower} negative. Since the choice of value of λ must vary continuously (using Lemma 143 and continuity of N^L), $\lambda = 0$ must be chosen at some point, i.e. $N^L(K_{\vec{y},\vec{z}}) = \vec{X}$, which is a contradiction.

Similarly a contradiction arises if when $\theta = 0$, $\lambda_{lower}(0)$ is chosen.

Case 2 For some $K = K_{\vec{y},\vec{z}}$, $N^L(K)$ is strictly positive, i.e. in the interior of $\mathbb{D}^3$.

In this case, let K_0 be such a K and w.l.o.g., we may assume that there is only one tool t such that

$$V^L(K) \cap t = \{N^L(K)\} \tag{7.75}$$

If not, note that there are only finitely many points where different tools intersect. Thus by choosing $\vec{y'}$ and $\vec{z'}$ close enough to $\vec{y}$, $\vec{z}$ respectively, $V^L(K_{\vec{y'},\vec{z'}})$ is as close as we want to $V^L(K_{\vec{y},\vec{z}})$ such that none of those intersections lie on it. Since N^L is continuous, we may assume that $N^L(K_{\vec{y'},\vec{z'}})$ is strictly positive.

Now we look at $V^L(K_0)$ in the form $Line(\vec{X}, \theta_0)$ of (7.72) where $X = N^L(K_0)$ is fixed and θ takes the value $\theta = \theta_0$. We assume that t is the only tool in the toolbox allowing N^L for which

$$t \cap Line(\vec{X}, \theta_0) = \{\vec{X}\} \tag{7.76}$$

so every other tool is bounded away from $\vec{X}$, say at least δ away.

We let a function of real-valued θ be given by

$$Nrotate(\theta) = N^L(K(\vec{X}, \theta)) \tag{7.77}$$

where $K(\vec{X}, \theta)$ is a knowledge base with solution set $Line(\vec{X}, \theta)$. By Lemma 143 this is a continuous function and $Nrotate(\theta_0) = \vec{X}$. However the quantity

$$|Nrotate(\theta) - \vec{X}| \tag{7.78}$$

also varies continuously, always taking the value zero or at least δ. By the Intermediate Value Theorem it must be constant at zero. Hence $N^L(t) = \vec{X}$, which is a contradiction since t is the only tool passing through $\vec{X}$.

We have proved the theorem in both cases.

■

Chapter 8

Introducing $Meanimax^L$, an interesting counterexample

In the book [Par] many Language Invariant inference processes are mentioned of the form

$$N(K) = \text{ the unique } \vec{x} \text{ for which } \sum_{i=1}^{J} f(x_i) \text{ is minimal} \tag{8.1}$$

where the function $f : [0, 1] \to \mathbb{R}$ is independent of J.

Theorem 144 *If N satisfies the conditions above, N satisfies Atomic Renaming, Obstinacy, Relativisation and Irrelevant Certainty.*

Proof Atomic Renaming is a property of N because the definition is symmetrical w.r.t. any permutation of the variables x_i.

Obstinacy holds because, when the overlying language is fixed, the inference process chooses the solution of the knowledge base which is minimal w.r.t. a fixed partial ordering - recall Theorem 9.

We may generalise the proof of that used in [Par] to prove that *Maximum Entropy* satisfies Relativisation, to show that Relativisation is satisfied by all inference processes N of the form (8.1). For Irrelevant Certainty, we are done by Theorem 129. We have proved the theorem.

$\blacksquare$

Remark It might then seem reasonable for us to suppose that Irrelevant Certainty is *always* a consequence of Language Invariance, Atomic Renaming, Obstinacy and Relativisation.

However the inference process $Meanimax^L$ which we define in this chapter provides a counterexample to any such conjecture, even if we limit our attention to Piecewise Linear Loaf Continuous processes.

Notation We shall consider the overlying language L to be fixed, so J is known. Then $dis(x)$ denotes $|x - \frac{1}{J}|$ for all real numbers x.

Definition For all $\vec{x} \in \mathbb{R}^J$ define $\tilde{\vec{x}}$ to be the unique vector $\tilde{\vec{x}}$ which is a permutation of $(dis(x_1), \ldots dis(x_J))$ such that $\tilde{x}_1 \geq \tilde{x}_2 \ldots \geq \tilde{x}_J$. Then the inference process $Meanimax^L$ (where $J = |At^L|$) is given by:

$$Meanimax^L(K) = \text{ the unique } \tilde{\vec{x}} \text{ which is minimal in the lexicographic ordering} \tag{8.2}$$

and it is abbreviated to $Meanx^L$. We will show below that this is well-defined. Just as we did for the cases of $Minimax$ and $Maximin$, we define the *meanimax* ordering to be the partial ordering of vectors in $\mathbb{R}^J$ such that $\vec{x}$ is *meanimax*-better than $\vec{y}$ iff $\tilde{\vec{x}}$ is before $\tilde{\vec{y}}$ lexicographically. Thus $Meanx^L(K)$ is the *meanimax*-best solution of K.

Definition In this chapter, $d_1, d_2 \ldots d_J$ are functions of $K \in CL$ given by:

$$d_1 = min\{max_{1 \leq i \leq J} dis(x_i) \,|\, \vec{x} \in V^L(K)\} \tag{8.3}$$

and then

$d_{k+1} = min\{max\ dis(x_i) \,|$ there exist $i_1, i_2, \ldots i_k$ distinct from i and from each other and s.t. $dis(x_{i_1}), dis(x_{i_2}) \ldots dis(x_{i_k})$ are equal to $d_1, d_2, \ldots d_k$ respectively where $\vec{x} \in V^L(K)\}$.

Then we can see that $Meanx^L(K)$ is a vector $\vec{X}$ in $\vec{V_L}(K)$ such that $(dis(X_1), dis(X_2), \ldots dis(X_J))$ is a permutation of $(d_1, d_2, \ldots d_J)$. In other words,

$$\widetilde{Meanx^L}(K) = (d_1, d_2, \ldots d_J) \tag{8.4}$$

In this chapter, we use the notation $\tilde{}$ in the $Meanimax^L$ sense, unless otherwise stated.

$Meanimax^L$ minimises the maximum distance of an atomic belief from the average, $1/J$, and then minimises the maximum distance of the others from $1/J$ etc.. As with $Minimax$, $Maximin$ etc. the minima required exist at every stage because $V^L(K)$ is a compact subset of $\mathbb{D}^J$. Now $d_1 \geq d_2 \geq \ldots \geq d_J$ and we can see this in a similar way to the proof of Lemma 27. d_1 is the smallest possible largest value of $dis(x_i)$ when $\vec{x} \in V^L(K)$ and then if such an $\vec{x}$ satisfies $\tilde{x}_1 = d_1$, the smallest possible largest value of $dis(x_i)$, when one occurrence of d_1 is ignored, is d_2 so $d_1 \geq d_2$ etc..

Theorem 145 *Given $K \in CL$, there exists a bijection $\sigma : \{1, 2, \ldots J\} \to \{1, 2, \ldots J\}$ and $\vec{side} \in \{0, 1\}^J$ such that for each $k = 1, 2, \ldots J$ and all $\vec{x} \in V^L(K)$ s.t. $\tilde{x}_1 = d_1, \ldots \tilde{x}_k = d_k$,*

$$x_{\sigma(i)} = \frac{1}{J} + (-1)^{side_i} d_i \tag{8.5}$$

for each $i = 1, 2, \ldots k$.

Definition If, given K, the identity permutation can fulfil the role of σ above, we say that K *admits the identity permutation w.r.t. Theorem* 145.

Remark Although we shall specify w.l.o.g. that σ is the identity permutation, as we have done in the cases of $Maximin$ and $Minimax$, we do not specify w.l.o.g. a particular value of $\vec{side}$.

Essentially given that the furthest k atomic beliefs from $\frac{1}{J}$ (allowing for duplicates in the same sense as "k largest" was used for $Minimax$) are $d_1, d_2, \ldots d_k$ away not only can we fix k distinct co-ordinates whose values must be $\frac{1}{J} \pm d_i$ but also whether their values are above or below $\frac{1}{J}$.

Proof of Theorem 145 We do this by induction on $k = 1, \ldots J$.

Base Case <u>$k = 1$</u> Suppose that no such pair $(\sigma(1), side_1)$ exists. Then for every $i = 1, \ldots J$ and $s = 0$ or 1, there exists $\vec{X}^{(i,s)} \in V^L(K)$ s.t. $\tilde{X}_1^{(i,s)} = d_1$ and

$$X_i^{(i,s)} \neq \frac{1}{J} + (-1)^s d_1 \tag{8.6}$$

otherwise $\sigma(1) = i, side_1 = s$ would work. Now, in a similar way to the proof of Theorem 28, taking the average gives a contradiction, indeed if

$$\vec{y} = \frac{1}{2J}\sum_{i=1}^{J}(\vec{X}^{(i,0)} + \vec{X}^{(i,1)}) \tag{8.7}$$

then by convexity of $V^L(K)$, $\vec{y} \in V^L(K)$. Since every $\tilde{X}_1^{(i,s)} = d_1$,

$$X_j^{(i,s)} \in I_1 = \left[\frac{1}{J} - d_1, \frac{1}{J} + d_1\right] \tag{8.8}$$

for every $i, j \in \{1, \ldots J\}$, for $s = 0$ and $s = 1$. By convexity of I_1, all of the values of $\vec{y}$ are in I_1, so $\tilde{y}_1 \leq d_1$. Since d_1 is the minimum possible value of $\tilde{x}_1$ when $\vec{x} \in V^L(K)$, $\tilde{y}_1 = d_1$ so for some j' such that $1 \leq j' \leq J$ and $s' = 0$ or 1, $y_{j'} = \frac{1}{J} + (-1)^{s'}d_1$. Suppose that $s' = 0$. Then

$$2Jy_{j'} = \sum_{i=1}^{J} X_{j'}^{(i,0)} + \sum_{i=1}^{J} X_{j'}^{(i,1)} \leq X_{j'}^{(j',0)} + (2J-1)\left(\frac{1}{J} + d_1\right) \tag{8.9}$$

but $X_{j'}^{(j',0)} < \frac{1}{J} + d_1$ by definition so $2Jy_{j'} < 2J\left(\frac{1}{J} + d_1\right)$ and $y_{j'} < (\frac{1}{J} + d_1)$, which is a contradiction. We get a similar contradiction if $s' = 1$, completing the Base Case.

Inductive Step We assume (I.H.) that we have fixed distinct $\sigma(1) \ldots \sigma(k)$ and $side_1 \ldots side_k$ such that for all solutions $\vec{x}$ of K s.t. $\tilde{x}_i = d_i$ for each $i = 1, \ldots k$,

$$x_{\sigma(i)} = \frac{1}{J} + (-1)^{side_i}d_i \tag{8.10}$$

for those values of i.

Assume for contradiction that there do not exist $\sigma(k+1)$ distinct from $\sigma(1) \ldots \sigma(k)$, together with $side_{k+1} \in \{0, 1\}$ such that for every $\vec{x} \in V^L(K)$ s.t. $\tilde{x}_1 = d_1 \ldots\ldots$ $\tilde{x}_{k+1} = d_{k+1}$, $x_{\sigma(i)} = \frac{1}{J} + (-1)^{side_i}d_i$ for every i s.t. $i \leq k+1$. Let

$$Others = \{1 \ldots J\} \setminus \{\sigma(1), \ldots \sigma(k)\} \tag{8.11}$$

Then we let $\vec{x} \in V^L(K)$ be s.t. $x_{\sigma(i)} = \frac{1}{J} + (-1)^{side_i}d_i$ for every $i = 1 \ldots k$ and suppose that the largest $dis(x_g)$ for which $g \in Others$ is d_{k+1}. Just as for the Base Case, for every $g \in Others$ $\vec{X}^{(g,s)}$ exist whose values at all c'th co-ordinates, where $c \in Others$, are in the interval

$$I_{k+1} = \left[\frac{1}{J} - d_{k+1}, \frac{1}{J} + d_{k+1}\right] \tag{8.12}$$

If $\vec{y}$ is the average of these $\vec{X}^{(g,s)}$, then $y_{\sigma(i)} = \frac{1}{J} + (-1)^{side_i} d_i$ for $i = 1, \dots k$; but if the next furthest value $y_{j'}$ from $\frac{1}{J}$ is $\frac{1}{J} + (-1)^{s'} d_{k+1}$ then a contradiction follows since $X_{j'}^{(j',0)} < \frac{1}{J} + d_{k+1}$ if $s' = 0$, while $X_{j'}^{(j',1)} > \frac{1}{J} - d_{k+1}$ implies a contradiction when $s' = 1$. Thus we have proved the Inductive Step and the theorem.

■

Corollary 146 *$Meanimax^L$ is well-defined.*

Proof The case $k = J$ in the above theorem implies the existence of a permutation $\sigma : \{1, \dots J\} \to \{1, \dots J\}$ together with $\vec{side} \in \{0,1\}^J$ such that if $\vec{x} \in V^L(K)$ and $\tilde{\vec{x}} = (d_1, \dots d_J)$,

$$x_i = \frac{1}{J} + (-1)^{side_{\sigma^{-1}(i)}} d_{\sigma^{-1}(i)} \tag{8.13}$$

for $i = 1, \dots J$. If $\vec{X} = Meanimax^L(K)$ we know that $\vec{X} \in V^L(K)$ and $\tilde{\vec{X}} = (d_1, \dots d_J)$ so the value of $\vec{X}$ is forced and we have proved the corollary.

■

Lemma 147 *$Meanimax^L$ satisfies Equivalence, Atomic Renaming and Obstinacy.*

Proof Equivalence is satisfied because the definition is given in terms of $V^L(K)$. Atomic Renaming is trivial because the definition is symmetrical w.r.t. the coordinates $1, \dots J$. Obstinacy is satisfied because $Meanimax^L$ chooses the optimal $\vec{x}$ w.r.t. a fixed partial ordering. Hence, by Theorem 9, we have proved the lemma.

■

The following lemma will be useful for us to prove that $Meanimax^L$ satisfies both Language Invariance and Relativisation.

Lemma 148 *Let $C \subset \{1, \dots J\}$ and let $\sim$ be an equivalence relation on $\{1 \dots J\} \setminus C$ such that the equivalence classes are all of equal size. Let $i_1, \dots i_q$ be representatives of the q distinct equivalence classes. For each $\vec{x} \in \mathbb{D}^J$ s.t. $x_i = x_j$ for all i, j such that $i \sim j$, let $Simp(\vec{x}) = (x_{i_1} \dots x_{i_q})$. Then if $\vec{y} \in \mathbb{R}^J$ is s.t. $y_c = x_c$ for all $c \in C$, comparing $\vec{x}$ and $\vec{y}$ in the meanimax ordering is equivalent to comparing $Simp(\vec{x})$ and $Simp(\vec{y})$ in the meanimax ordering.*

Corollary 149 *Suppose $\vec{x}, \vec{y}$ are vectors in $\mathbb{R}^J$ such that $x_i = y_i$ for all $i \in C$, for some $C \subset \{1, \ldots J\}$. W.l.o.g. let $C = \{1, \ldots k\}$. We can do this by the symmetry of the meanimax ordering w.r.t. permuting the co-ordinates. Then comparing $\vec{x}$ and $\vec{y}$ in the meanimax ordering is equivalent to comparing $(x_{k+1}, \ldots x_J)$ and $(y_{k+1}, \ldots y_J)$. If the same C is a subset of co-ordinates which are constant w.r.t. a knowledge base $K \in CL$, then $Meanimax^L(K) =$ that $\vec{x} \in V^L(K)$ for which $(x_{k+1}, \ldots x_J)$ is meanimax-best.*

Proof of Lemma 148 Comparing $\vec{x}$ and $\vec{y}$ in the *meanimax* ordering is equivalent to comparing $dis(\vec{x})$ and $dis(\vec{y})$ in the *minimax* ordering, where

$$dis(\vec{x}) = (dis(x_1), dis(x_2), \ldots dis(x_J)) \tag{8.14}$$

by definition, for all $\vec{x}$. Thus, for every $\vec{x}$,

$$dis(Simp(\vec{x})) = \left(\left| x_{i_1} - \frac{1}{J} \right|, \left| x_{i_2} - \frac{1}{J} \right|, \ldots \left| x_{i_q} - \frac{1}{J} \right| \right) = Simp(dis(\vec{x})) \tag{8.15}$$

Hence for $\vec{x}$, $\vec{y}$ as in the lemma, $\vec{x}$ is *meanimax*-better than $\vec{y}$

$\Leftrightarrow$ $dis(\vec{x})$ is *minimax*-better than $dis(\vec{y})$

$\Longleftrightarrow$ $Simp(dis(\vec{x}))$ is *minimax*-better than $Simp(dis(\vec{y}))$ (see Lemma 42)

$\Longleftrightarrow$ $dis(Simp(\vec{x}))$ is *minimax*-better than $dis(Simp(\vec{y}))$

$\Longleftrightarrow$ $Simp(\vec{x})$ is *meanimax*-better than $Simp(\vec{y})$.

Hence we have proved the lemma.

■

Proof of Corollary 149 This is essentially the same as the proof of Corollary 43.

■

Lemma 150 *$Meanimax^L$ is Language Invariant.*

Proof This proof is much the same as the corresponding proof for *Maximin* - see Theorem 73. Indeed suppose that $At^L = \{\alpha_1, \ldots \alpha_J\}$ and there exists $K \in CL$ s.t. $Meanx^{L'}(K)(\alpha_i) \neq Meanx^L(K)(\alpha_i)$ for some $i = 1, \ldots, J$ and $L' = L + \{p'\}$, where $p' \notin L$. For each $\alpha_i \in At^L$, we let $\beta_{(i,+)} = \alpha_i \wedge p'$ and $\beta_{(i,-)} = \alpha_i \wedge \neg p'$. By Atomic Renaming

$$\vec{W} \in \mathbb{D}^{2J} = Meanx^{L'}(K) \tag{8.16}$$

gives the same belief value to $\beta_{(i,+)}$ and to $\beta_{(i,-)}$ for each $j = 1 \ldots J$. The proof now continues as for $Maximin$ to get a contradiction but using *meanimax* comparisons and Lemma 148. We have proved the lemma.

■

From now on we usually write $Meanimax$, $Meanx(K)$ etc. without mentioning the overlying language.

Lemma 151 *$Meanimax$ satisfies Relativisation.*

Proof To prove this we can use exactly the same method as in the proof of Theorem 83, but exchanging $Maximin$ and *maximin* for $Meanimax$ and *meanimax* respectively and using Corollary 149 instead of Corollary 75.

■

Definition For any language L, we let

$$T^L_{Mean} = \left\{ (x_i = x_j), (x_i = 0), \left(x_i + x_j = \frac{2}{J} \right) \text{ s.t. } 1 \leq i < j \leq J \right\} \tag{8.17}$$

and we say that an *EZM-constraint* is any constraint which is either an EZ-constraint or is of the form $x_i + x_j = \frac{2}{J}$.

Theorem 152 *$Meanimax$ is PL, and for every language L, $Meanimax^L$ is allowed by the toolbox T^L_{Mean}.*

Notation In the following proof we use constraint sets instead of knowledge bases, in a manner similar to Sections 6.1 and 6.2. The work so far in this chapter goes through for constraint sets, *apart from the fact that the number of co-ordinates of a constraint set does not necessarily equal the J which is the number of co-ordinates of vectors in the meanimax ordering we are using, the values of dis, EZM-constraints etc.* Where necessary, we denote the number of co-ordinates of a constraint set K

by J_K. However, J *always* denotes the number of co-ordinates of vectors in the *meanimax* ordering we use.

Proof We start by defining the following operation on constraint sets:

Definition Let $K \in \mathcal{C}^d_{J_K}$, s.t. K has two distinct solutions. The equivalence relation K-*mean equivalence* on $\{1, 2, \ldots J_K\} \backslash \{K\text{-constant co-ordinates}\}$, denoted by $\sim_K$, is given by

$$i \sim_K j \iff \left(x_i = x_j \text{ or } x_i + x_j = \frac{2}{J} \right) \text{ for all } \vec{x} \in V(K) \tag{8.18}$$

where the equivalence classes w.r.t. $\sim_K$ are given by $[e_1], \ldots [e_Q]$ and $Q \geq 2$ since K has more than one solution.

Note that if, for non-constant co-ordinates i, j, $x_i = x_j$ or $x_i + x_j = \frac{2}{J}$ for every $\vec{x} \in V(K)$ then either $x_i = x_j$ for every solution or $x_i + x_j = \frac{2}{J}$ for every solution since otherwise there exists a solution $\vec{X}^{(1)}$ s.t. $X^{(1)}_i = X^{(1)}_j$ but $X^{(1)}_i + X^{(1)}_j \neq \frac{2}{J}$ and a solution $\vec{X}^{(2)}$ s.t. $X^{(2)}_i \neq X^{(2)}_j$ and $X^{(2)}_i + X^{(2)}_j = \frac{2}{J}$. The average

$$\vec{z} = \frac{\vec{X}^{(1)} + \vec{X}^{(2)}}{2}, \text{ for which } z_i \neq z_j, z_i + z_j \neq \frac{2}{J} \tag{8.19}$$

is also a solution, contradicting our assumptions.

For every K-constant co-ordinate c_i, let C_i be the value such that $x_{c_i} = C_i$ for all $\vec{x} \in V(K)$.

Definition We now define the $\underline{Meanimax\text{-simplification of K}}$, denoted by $MeanS(K)$, as follows:

If c_i is a constant co-ordinate, we substitute C_i for each appearance of x_{c_i} in the constraints of K. We now consider, in turn, each non-constant co-ordinate q. If, for some $p = 1, \ldots Q$, $q \sim_K e_p$, and $x_{e_p} = x_q$ for every $\vec{x} \in V(K)$, we substitute y_p for each appearance of x_q in K. However, if $q \sim_K e_p$ and $x_{e_p} + x_q = \frac{2}{J}$ for all $\vec{x} \in V(K)$ we substitute $\frac{2}{J} - y_p$ for each appearance of x_q in K. The resulting set of equations on $y_1, \ldots y_Q$ is $MeanS(K)$. The notation y is chosen just so as to be distinct from x. Define

$$MeanS_K : V(K) \to V(MeanS(K)) \text{ given by } MeanS_K(\vec{x}) = (x_{e_1}, \ldots x_{e_Q}) \tag{8.20}$$

Lemma 153 *$MeanS_K : V(K) \to V(MeanS(K))$ is well-defined and is a bijection whose inverse we denote by $MeanS_K^{-1}$.*

Proof If $\vec{x}$ satisfies K, by checking the constraints individually we see that $MeanS_K(\vec{x})$ satisfies $MeanS(K)$. Non-negativity of $MeanS_K(\vec{x})$ is also clear.

If $\vec{y}$ satisfies $MeanS(K)$, we can find a unique $\vec{X}$ such that $MeanS_K(\vec{X}) = \vec{y}$ as follows: Consider each co-ordinate $i = 1, \dots J_K$ in turn. If i is K–constant then $i = c_j$ for some j so we are forced to let $X_i = C_j$. Otherwise, there exists a unique $p = 1 \dots Q$ such that $i \sim_K e_p$. If $x_{e_p} = x_i$ for all $\vec{x} \in V(K)$ we must let $X_i = y_p$. Otherwise, $x_{e_p} = \frac{2}{J} - x_i$ for all $\vec{x} \in V(K)$ so we must let $X_i = \frac{2}{J} - y_p$. There is no choice at any step in this procedure, which does find such an $\vec{X}$.

Hence $MeanS_K$ is a bijection and we know how to calculate its inverse. We have proved the lemma.

■

Lemma 154 *If $K \in \mathcal{C}^d_{J_K}$, K has more than one solution and $\sim_K$, $e_1 \dots e_Q$ etc. as above, $MeanS_K(Meanx(K)) = Meanx(MeanS(K))$.*

Let $Meanx(K) = \vec{X}$ and $Meanx(MeanS(K)) = \vec{Y}$, where $\vec{a} = MeanS_K(\vec{X}) \in \mathbb{R}^Q$ and $\vec{b} = MeanS_K^{-1}(\vec{Y})$. We assume w.l.o.g. that $MeanS(K)$ admits the identity permutation w.r.t. Theorem 145.

We show by induction on p that

Claim(p) $\tilde{a}_1 = \tilde{Y}_1 \dots \tilde{a}_p = \tilde{Y}_p$, so that $Y_1 = a_1 \dots Y_p = a_p$.

Base Case $\underline{p = 0}$ In this case the claim is trivial.

Inductive Step Given that the largest p values in $dis(\vec{a})$ match the largest p values in $dis(\vec{Y})$, $\tilde{X}_k = \tilde{b}_k$ for each k not greater than

$$FIX = \left(\sum_{i=1}^{p} |[e_i]| \right) + |\{j \mid C_j \geq a_p\}| \tag{8.21}$$

by the procedure for calculating $MeanS_K^{-1}$. Suppose for contradiction that $\tilde{Y}_{p+1} \neq \tilde{a}_{p+1}$. Then, in a similar way to the proof of Lemma 96, $\tilde{b}_{FIX+1} < \tilde{X}_{FIX+1}$, contradicting the fact that $\vec{X} = Meanx(K)$. We have proved the lemma. ■

Lemma 155 *Let $K \in \mathcal{C}^d_{J_K}$ such that K has no constant co-ordinates and the relation $\sim_K$ is equality. Then if $\vec{X} = Meanx(K)$, either $X_i = 0$ for some i or $X_i = X_j$ or $X_i + X_j = \frac{2}{J}$ for some i, j s.t. $i \neq j$.*

Proof Suppose that K is some constraint set satisfying the hypotheses of the lemma but that all three possible conclusions above fail. Then, assuming w.l.o.g. that K admits the identity permutation w.r.t. Theorem 145, we have that

$$dis(X_1) > dis(X_2) > \ldots > dis(X_{J_K}) \tag{8.22}$$

since $dis(X_i) = dis(X_j)$ implies that $X_i = X_j$ or $X_i + X_j = \frac{2}{J}$. The rest of this proof is similar to that of Lemma 97. Since the co-ordinate 1 is not K-constant, there exists $\vec{u} \in \mathbb{R}^{J_K}$ parallel to $G(K)$ such that $u_1 > 0$ and

$$\{\vec{z} = \vec{X} + y\vec{u} \,|\, y \in \mathbb{R}, \vec{z} \text{ strictly positive, and } dis(z_1) > dis(z_i) \text{ for } i = 2 \ldots J_K\} \tag{8.23}$$

is an open set so there exists such a $\vec{z}$ s.t. $\tilde{z}_1 = dis(z_1) < dis(X_1) = d_1$, which is a contradiction. We have proved the lemma.

■

Lemma 156 *For any $K \in \mathcal{C}^d_{J_K}$ with more than one solution, there does not exist a $MeanS(K)$-constant co-ordinate, nor do there exist distinct i, j such that $y_i = y_j$ for all $\vec{y} \in V(MeanS(K))$. There do not exist distinct i, j such that $y_i + y_j = \frac{2}{J}$ for all $\vec{y} \in V(MeanS(K))$.*

Proof This is very similar to the proof of Lemma 98.

■

Lemma 157 *If $K \in \mathcal{C}^d_{J_K}$ and K has more than 1 solution, $Meanx(K)$ satisfies an EZM-constraint, which is not true throughout $V(K)$.*

Proof This is very similar to the proof of Lemma 99.

■

Proof of Theorem 152 The following lemma is essentially similar to Lemma 100.

Lemma 158 *Let $K \in \mathcal{C}^d_{J_K}$. Then for each $q = 0, \ldots d$ there exist q distinct EZM-constraints $E_1, \ldots E_q$, s.t. if K^+ denotes $K + E_1 + \ldots + E_q$ then*

$$Meanx(K) = Meanx(K^+) \tag{8.24}$$

and $rank(K^+) = J_K - d + q$.

Proof We do this by induction on q.

Base Case $\underline{q = 0}$ In this case the lemma is vacuous.

Inductive Step We assume (I.H.) that for some $q = k < d$, there exist EZM-constraints $E_1, \ldots E_k$ such that if we let

$$K^+ = K + E_1 + \ldots + E_k, \text{ then } rank(K^+) = J_K - d + k \tag{8.25}$$

and $Meanx(K) = Meanx(K^+) = \vec{X}$. If K^+ has a unique solution but $rank(K^+) < J_K$ then adding the constraint $(x_i = 0)$ for some i works in the same way as in the proof of Lemma 100. Otherwise Lemma 157 implies that we can fix an EZM-constraint E_{k+1} which is not true in all of $V^L(K^+)$, but which is satisfied by $\vec{X}$. Then, if

$$\hat{K^+} = K^+ + E_{k+1} \tag{8.26}$$

$rank(\hat{K^+}) = J_K - d + k + 1$ and $Meanx(\hat{K^+})$ is a solution of (K^+) so by Obstinacy (Lemma 147 for constraint sets), $\vec{X} = Meanx(\hat{K^+})$, completing the Inductive Step and the proof of the lemma.

■

Proof of Theorem 152 If we let $q = d$ in Lemma 158, then we have proved the theorem.

■

Remark Thus we can use an algorithm which is an analogue of Algorithm 94 to calculate *Meanimax*, say the *Meanimax* Calculation Algorithm. It differs from Algorithm 94 only insofar as the constraints we add at Step 2 are EZM-constraints, and in Steps 4 and 5 we sort the members of *Candidates* using the *meanimax* ordering instead of the *minimax* ordering.

Notation For the remainder of this chapter we refer to knowledge bases, rather than constraint sets in general, so the notation J_K is not required.

Theorem 159 *$Meanimax$ is Piecewise Linear Loaf Continuous.*

Proof This is very similar to the corresponding proof for $Minimax$ - see Theorem 51. As before we prove w.l.o.g., for a loaf of the form

$$\hat{K} = \{K_\lambda \,|\, \lambda \in [0,1]\} \tag{8.27}$$

that $Meanx(K_\lambda) \to 0$ as $\lambda \searrow 0$.

Lemma 160 *The function $d_1 : CL \to \mathbb{R}$ given by*

$$d_1(K) = min_{\vec{x} \in V^L(K)} max_{1 \le i \le J} dis(x_i) = \widetilde{Meanx(K)}_1 \tag{8.28}$$

is uniformly continuous (using the Blaschke topology).

Proof Suppose $K_1, K_2 \in CL$ and $\Delta(K_1, K_2) < \epsilon$, where $\epsilon > 0$.
Then if $\vec{x} = Meanx(K_1)$ there exists $\vec{y} \in V^L(K_2)$ s.t. $|\vec{x} - \vec{y}| < \epsilon$, so $|\tilde{x}_1 - \tilde{y}_1| < \epsilon$, and as $\tilde{x}_1 = d_1(K_1)$ we see that

$$\tilde{y}_1 < d_1(K_1) + \epsilon \text{ so } d_1(K_2) < d_1(K_1) + \epsilon \tag{8.29}$$

Similarly $d_1(K_1) < d_1(K_2) + \epsilon$ so d_1 is uniformly continuous.

■

Proof of Theorem 159 The rest of this closely follows the proof of Theorem 51. Thus it is true that when we use the $Meanimax$ Calculation Algorithm (as discussed above) instead of Algorithm 94, we obtain constraint sets, the left hand sides of which are independent of λ. Whether these constraint sets have unique solutions or not is independent of λ. We then obtain a finite list of linear functions of λ, as in (3.48), such that for all $\lambda \in [0,1]$, $Meanx(K_\lambda)$ must equal one of those functions. Comparing these possible values in the *meanimax* ordering gives the same linear formula for small enough positive λ, so

$$Meanx(K_\lambda) = \vec{p} + \lambda\vec{g} \tag{8.30}$$

for those small $\lambda > 0$, where $\vec{p} \in V^L(K_0)$.

We suppose for contradiction of continuity at $\lambda = 0$ that $Meanx(K_0) = \vec{b} \neq \vec{p}$ and assume w.l.o.g. that K_0 admits the identity permutation w.r.t. Theorem 145 with some $\vec{side}$ fixed to go with it so that $dis(b_1) \geq dis(b_2) \ldots \geq dis(b_J)$. The following lemma gives us the required contradiction.

Lemma 161 *For each $i = 1 \ldots J$, $\tilde{p}_i = dis(b_i)$ and $p_i = b_i$.*

Proof We do this by induction on i.

Base Case $\underline{i = 1}$ Now $lim_{\lambda \to 0}(\widetilde{\vec{p} + \lambda\vec{g}})_1 = \tilde{p}_1$, since $\tilde{}_1 = max(dis)$ is a continuous function and that limit equals $\tilde{b}_1$ because d_1 is continuous. Hence $\tilde{p}_1 = dis(b_1)$ so $p_1 = b_1$ because the choice between $\frac{1}{J} \pm dis(b_1)$ is fixed by $side_1$.

Inductive Step Assume (I.H.) that the lemma holds for $i = 1, \ldots k$. Suppose for contradiction that $\tilde{p}_{k+1} \neq \tilde{b}_{k+1}$, then

$$\tilde{p}_{k+1} > \tilde{b}_{k+1} = d_{k+1} \tag{8.31}$$

Similarly to the proof of Lemma 52, since all of the $G^L(K_\lambda)$ are parallel, $\vec{b} + \lambda\vec{g} \in G^L(K_\lambda)$ for $\lambda \in [0, 1]$ and, for small enough non-negative λ,

$$\vec{y(\lambda)} = \vec{p} + \frac{1}{2}(\vec{b} - \vec{p}) + \lambda\vec{g} \tag{8.32}$$

is non-negative and so in $V^L(K_\lambda)$. Then

$$w(\vec{\alpha}, \lambda) = \vec{p} + \alpha(\vec{b} - \vec{p}) + \lambda\vec{g} \in V^L(K_\lambda) \tag{8.33}$$

for those small λ and every $\alpha \in [0, 1/2]$. Let $\vec{z} = \vec{p} + \lambda\vec{g}$. Then for each $i = 1, \ldots k$, $w_i = z_i$. Now $\vec{w}$ is *meanimax*-better than $\vec{z}$ iff $dis(\vec{w})$ is *minimax*-better than $dis(\vec{z})$ iff $dis(\vec{W})$ is *minimax*-better than $dis(\vec{Z})$, where

$$\vec{W} = (W_{k+1}, W_{k+2} \ldots W_J) \text{ and } \vec{Z} = (Z_{k+1}, \ldots Z_J) \tag{8.34}$$

by Corollary 43. Hence $\vec{w}$ is *meanimax*-better than $\vec{z}$ iff $\vec{W}$ is *meanimax*-better than $\vec{Z}$.

If $\lambda = 0$ and $\alpha = 0$, $\tilde{Z}_1 > \tilde{W}_1$ but by continuity of the function $\tilde{\ }_1$ this remains true for small values of λ and α, contradicting the known fact that $\vec{z} = Meanx(K_\lambda)$. Hence

$$\tilde{p}_{k+1} = \tilde{b}_{k+1} = d_{k+1} \tag{8.35}$$

and by using σ, $side_{k+1}$, (by Lemma 145) we see that $p_{k+1} = b_{k+1}$, completing the Inductive Step. We have proved the lemma.

■

Proof of Theorem 159 continued The theorem now follows in just the same way as for Theorem 51.

■

Theorem 162 *Meanimax does not satisfy Irrelevant Certainty.*

Remark Intuitively this is because when we add an extra propositional variable p' and the constraint $Bel(p') = 0$, although the constraints look the same on $x_1, \ldots x_J$ with constraints $x_{J+1} = \ldots = x_{2J} = 0$ added, the largest value of $|x_i - \frac{1}{2J}|$ is minimised instead of the maximal value of $|x_i - \frac{1}{J}|$.

Proof Let

$$K_1 = \{Bel(p_1 \wedge \neg p_2) = 2Bel(p_1 \wedge p_2), Bel(p_2 \wedge \neg p_1) = 3Bel(p_1 \wedge p_2)\} \tag{8.36}$$

Then, using the standard ordering of the atoms

$$V^L(K) = \{(\tau, 2\tau, 3\tau, 1 - 6\tau) \text{ s.t. } 0 \leq \tau \leq 1/6\} \tag{8.37}$$

We let $\vec{X} = Meanx(K)$ and show that $\vec{X} = \left(\frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{2}{8}\right)$. $\tilde{X}_1 = 1/8$ since if $\tau < 1/8$, $\tilde{x}_1 > 1/8$ for the resulting solution $\vec{x}$ but $\tau > 1/8$ gives $x_3 > 3/8$ so $\tilde{x}_1 > 1/8$. Let

$$K_2 = K_1 + Bel(p') = 0 \tag{8.38}$$

then using the standard ordering of the atoms of $L' = L + p'$,

$$V^{L'}(K_2) = \{(\tau, 2\tau, 3\tau, 1 - 6\tau, 0, 0, 0, 0) \text{ s.t. } 0 \leq \tau \leq 1/6\} \tag{8.39}$$

We let $Meanx(K_2) = \vec{Y}$ and show that $\vec{Y} = \left(\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{3}{9}, 0, 0, 0, 0\right)$. $\tilde{Y}_1 = 5/24$ since $\tau < 1/9$ implies that $x_4 > 1/3$ for the resulting solution $\vec{x}$ but $\tau > 1/9$ gives $x_3 > 1/3$ so either way $\tilde{x}_1$ is greater than $5/24$.

Thus Irrelevant Certainty fails since $Meanx(K_1)(p_1) = 3/8$ but $Meanx(K_2)(p_1) = 1/3$ so we have proved the theorem.

■

Theorem 163 *Meanimax does not satisfy Homogeneity.*

Proof For all $\lambda \in (0, 1]$, let K_λ be given by

$$K_\lambda = \left\{ Bel(p_1 \vee p_2) = \lambda, Bel(\neg p_2 | p_1) = \frac{2}{3} \right\} \tag{8.40}$$

When we use the overlying language $L = \{p_1, p_2\}$ and the standard ordering of the atoms of L,

$$V^L(K_\lambda) = \{(\tau, 2\tau, \lambda - 3\tau, 1 - \lambda) \text{ s.t. } 0 \le \tau \le \lambda/3\} \tag{8.41}$$

Let $\vec{X} = Meanx(K_1)$. We show that $\vec{X} = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}, 0\right)$. Now $max_{i=1,2,3,4}|X_i - 1/4| = 3/20$ since $\tau > 1/5$ gives $2\tau > 2/5$ and $\tau < 1/5$ gives $1 - 3\tau > 2/5$.

If Homogeneity is a property of $Meanimax$,

$$Meanx(K_\lambda) = \left(\frac{\lambda}{5}, \frac{2\lambda}{5}, \frac{2\lambda}{5}, 1 - \lambda \right) \tag{8.42}$$

for all $\lambda \in (0, 1]$ but we now show that $Meanx(K_{1/10}) = \left(\frac{1}{40}, \frac{2}{40}, \frac{1}{40}, \frac{36}{40}\right)$. For each solution of $K_{1/10}$, $\vec{x}$, $\tilde{x}_1 = 26/40$ since $x_4 = 36/40$. Also if $\tilde{x}_2 < \frac{9}{40}$ then $x_1 > \frac{1}{40}$ and $x_3 > \frac{1}{40}$. However $\tau > \frac{1}{40}$ gives $\frac{1}{10} - 3\tau < \frac{1}{40}$. Hence, since $\tilde{X}_2 = \frac{9}{40}$, $Meanx(K_{1/10}) = \left(\frac{1}{40}, \frac{2}{40}, \frac{1}{40}, \frac{36}{40}\right)$ and we have found a contradiction, so we have proved the theorem.

■

Remark This means that in Theorem 130 if Irrelevant Certainty is removed from the hypotheses, we cannot deduce Homogeneity.

We now, for completeness, examine how $Meanimax$ fares when tested against all of the Par-Ven Properties.

Theorem 164 *$Meanimax$ satisfies Equivalence, Atomic Renaming, Obstinacy, Language Invariance and Relativisation. However, it fails to satisfy Continuity, Open-mindedness, Independence and Irrelevant Information.*

Proof Lemmas 147, 150 and 151 give us the five Par-Ven properties that $Meanimax$ satisfies. Since $Meanimax$ is Partly Linear (by Theorem 152) we use Theorem 141 to see that $Meanimax$ does not satisfy Independence, and Theorem 142 to show that it is not continuous. By Lemma 162, $Meanimax$ does not satisfy Irrelevant Information. The following lemma is now sufficient to establish the theorem:

Lemma 165 *$Meanimax$ does not satisfy Open-mindedness.*

Proof To prove this we define K just as in the proof of Theorem 41, so $J = 4$. We see that

$$Meanx^L(K) = \left(0, 0, \frac{1}{2}, \frac{1}{2}\right) \tag{8.43}$$

This is because $x_3 = x_2 + 1/2$ for all $\vec{x} \in V^L(K)$, so $dis(x_3) \geq 1/4$. The only possible way that $\vec{x} \in V^L(K)$ and

$$max_{i=1,2,3,4} dis(x_i) = 1/4 \tag{8.44}$$

is if $dis(x_3) = 1/4$. This implies that $x_3 = 1/2$ so $x_2 = 0$, forcing (8.43).

However, similarly to the proof of Theorem 41, $x_2 = 0$ is not necessarily true when $\vec{x} \in V^L(K)$. Hence $Meanimax$ does not satisfy Open-mindedness and we have proved the lemma and Theorem 164.

■

Chapter 9

Conclusions

We have found that the inference process *Minimax* is the upper limit of the *Renyi Processes* and discovered its properties, tabulated below. We have also justified *Maximin* as the limit of another sequence of inference processes. We have found that *Maximin* compares favourably with *Minimax* when we consider the set of properties known as the Par-Ven Properties, due to [ParVen1], which allow us to uniquely characterise *Maximum Entropy*. Indeed, if judged by the set of desirable properties which it satisfies, the table overleaf shows that *Maximin* is perhaps the inference process which is the closest rival to *Maximum Entropy* which has hitherto been discovered.

In the table below, the properties abbreviated in the leftmost column are: Equivalence, Atomic Renaming, Obstinacy, Language Invariance, Continuity, Open-mindedness, Independence, Relativisation, Irrelevant Information, Piecewise Linear Loaf Continuity, Partial Linearity, Irrelevant Certainty and Homogeneity.

In the table, $*$ means the result is due to [ParVen1]. † means the result is due to [Moh]. If a result is subscripted with a number, that is the number of the lemma or theorem in this work of which it is a direct consequence. Anything in brackets is a conjecture.

	CM_∞	*Renyi Processes* Ren_r $(r > 0)$ $r < 1$	$r = 1$ (ME)	$r > 1$, $r \neq 2$	$r = 2$ (MD)	Mmx	$Mxmn$	$Meanx$
Equiv	$\checkmark_{21}$	$\checkmark_{\dagger}$	$\checkmark_{*}$	$\checkmark_{\dagger}$	$\checkmark_{\dagger}$	$\checkmark_{35}$	$\checkmark_{70}$	$\checkmark_{164}$
At Ren	$\checkmark_{21}$	$\checkmark_{\dagger}$	$\checkmark_{*}$	$\checkmark_{\dagger}$	$\checkmark_{\dagger}$	$\checkmark_{36}$	$\checkmark_{71}$	$\checkmark_{164}$
Obs	$\checkmark_{21}$	$\checkmark_{\dagger}$	$\checkmark_{*}$	$\checkmark_{\dagger}$	$\checkmark_{\dagger}$	$\checkmark_{37}$	$\checkmark_{72}$	$\checkmark_{164}$
Lan Inv	$\checkmark_{21}$	$\checkmark_{\dagger}$	$\checkmark_{*}$	$\checkmark_{\dagger}$	$\checkmark_{\dagger}$	$\checkmark_{39}$	$\checkmark_{73}$	$\checkmark_{164}$
Cont	$\times_{21}$	$\checkmark_{\dagger}$	$\checkmark_{*}$	$\checkmark_{\dagger}$	$\checkmark_{\dagger}$	$\times_{40}$	$\times_{77}$	$\times_{164}$
Open-min	$\checkmark_{21}$	$\checkmark_{\dagger}$	$\checkmark_{*}$	$\times_{\dagger}$	$\times_{\dagger}$	$\times_{41}$	$\checkmark_{78}$	$\times_{164}$
Indep	$\times_{21}$	$\times_{\dagger}$	$\checkmark_{*}$	$\times_{\dagger}$	$\times_{\dagger}$	$\times_{44}$	$\times_{79}$	$\times_{164}$
Relat	$\checkmark_{21}$	$\checkmark_{\dagger}$	$\checkmark_{*}$	$\checkmark_{\dagger}$	$\checkmark_{\dagger}$	$\checkmark_{48}$	$\checkmark_{83}$	$\checkmark_{164}$
Irrel Inf	$\times_{21}$	$(\times_{\dagger})$	$\checkmark_{*}$	$\times_{\dagger}$	$\times_{\dagger}$	$\times_{49}$	$\checkmark_{84}$	$\times_{164}$
PLLC	$\times_{53}$	$\times_{53}$	$\times_{53}$	$\times_{53}$	$\checkmark_{54}$	$\checkmark_{51}$	$\checkmark_{88}$	$\checkmark_{159}$
PL	$(\times)$	$\times_{142}$	$\times_{142}$	$\times_{142}$	$\times_{142}$	$\checkmark_{127}$	$\checkmark_{127}$	$\checkmark_{152}$
Irrel Cer	$\checkmark_{129}$	$\checkmark_{129}$	$\checkmark_{129}$	$\checkmark_{129}$	$\checkmark_{129}$	$\checkmark_{129}$	$\checkmark_{129}$	$\times_{162}$
Homog	$\checkmark_{134}$	$\checkmark_{133}$	$\checkmark_{133}$	$\checkmark_{133}$	$\checkmark_{133}$	$\checkmark_{132}$	$\checkmark_{132}$	$\times_{163}$

9.1 Unanswered questions for further study

There are many questions which arise naturally from the work in this thesis but which have not been explored.

- Are there other "natural" inference processes, apart from MD, which satisfy Piecewise Linear Loaf Continuity but which are not Partly Linear?
- Do there exist $K \in CL$ and $n \in \mathbb{N}$ such that n is not less than the number of possible outcomes consistent with K but $Every^L_n(K)$ is undefined?
- Just how bad are the worst cases when we run $Maximin$ Calculation Algorithm 3?
- Are there any useful weakenings of Irrelevant Information apart from Irrelevant Certainty?
- To what extent can the algorithms for calculating $Maximin$ be adapted for calculating $Minimax$?
- Is $Maximin$ the only PL inference process that can be uniquely characterised by a list of desiderata—what is the best possible strengthening of Theorem 138?

- Given a language invariant family N of PL inference processes and an infinite series of toolboxes $T_1, T_2, \ldots$ which allow the inference processes $N^{L_1}, N^{L_2}, \ldots$ respectively, where $L_i = \{p_1, \ldots p_i\}$ for each $i \in \mathbb{N}$, precisely what kind of "natural" way should there be of naming all of the toolboxes simultaneously [1]?

- Does there exist a useful PL inference process which does not have a uniquely minimal toolbox?

[1]This was raised on page 175

Bibliography

[Ber] Jacob Bernoulli *Ars Conjectandi*, Basileae : impensis Thurnisiorum fratrum, 1713.

[Chr] D Christensen *Clever Bookies and Coherent Beliefs*, The Philosophical Review, 1991.

[Court] Paul Courtney *PhD Thesis*, Manchester University, 1992.

[Egg] H.G. Egglestone *Convexity*, Cambridge University Press, Cambridge, 1969.

[Fad] D.K. Fadeev *Zum Begriff der Entropie ciner endlichen Wahrscheinlichkeitsschemas, Arbeiten zur Informationstheorie I*, Berlin, Deutscher Verlag der Wissenschaften, 1957, pp 85–90.

[deFin1] B. de Finetti *Sul significato soggetivo della probabilità*, Fundamenta Mathematicae, 1931, Vol.17, pp.298–329.

[deFin2] B. de Finetti *La prévision: ses lois logiques, ses sources subjectives*, Annales de l'Institute Henri Poincare, 1937, Vol.7, pp.1–68.

[HalKol] J.Y. Halpern, D. Koller *Representation Dependence in Probabilistic Inference*, IJCAI 1995, pp. 1853–1860.

[Jay] E.T. Jaynes *Information theory and statistical mechanics*, Statistical Physics, K. Ford, Benjamin, New York, 1963, p.181.

[Kap] J.N. Kapur *Measures of Information and Their Applications*, Wiley, New Delhi 1994.

[Kem] J.G. Kemeny *Fair bets and inductive probabilities*, Journal of Symbolic Logic, 1955, Vol.20, pp.1–28.

[LawWil] J. Lawry and G.M.Wilmers *An axoimatic approach to systems of prior distributions in inexact reasoning*, *Knowledge representation and reasoning under uncertainty*, *Lecture Notes in Computer Science*, Springer-Verlag, 1992.

[Lov] A.O.Lovejoy *The Great Chain of Being: A Study of the History of an Idea*, Harvard University Press, 1998.

[Mau] I. Maung *PhD Thesis* Manchester University, 1992.

[Moh] Ismail Mohamed *MSc Thesis* Manchester University, 1998.

[Par] J. B. Paris *The Uncertain Reasoner's Companion—A Mathematical Perspective* Cambridge Tracts in Theoretical Computer Science No.39, Cambridge University Press, 1994.

[ParVen1] J.B.Paris and A. Vencovska *A note on the inevitability of maximum entropy*, International Journal of Approximate Reasoning, 1990, Vol.4, No.3, pp 183–224.

[ParVen2] J.B.Paris and A. Vencovska *A method for updating that justifies minimum cross entropy*, International Journal of Approximate Reasoning, 1992, Vol. 7 pp 1–8.

[ParVen3] J.B.Paris and A. Vencovska *On the applicability of Maximum Entropy to inexact reasoning*, International Journal of Approximate Reasoning, 1989, Vol. 3, No.1, pp.1–34.

[ParVen4] J.B.Paris and A. Vencovska *In defence of the Maximum Entropy inference process*, International Journal of Approximate Reasoning, 1997.

[Ram] F.P. Ramsey *The foundations of mathematics and other logical essays*, London, 1931.

[Ren1] A. Renyi *On measures of entropy and information*, Proceedings of the 4th Berkeley Symposium in Mathematical Statistics, 1960.

[Ren2] A. Renyi *Wahrscheinlichkeitsrechnung*, Berlin: Deutscher Verlag der Wissenschaften, 1962.

[Sha] C.E. Shannon and W. Weaver *The mathematical theory of communication*, University of Illinois Press, Urbana, Illinois, 1964.

[Shi] A. Shimony *Coherence and the axioms of confirmation*, Journal of Symbolic Logic, 1955, Vol.20, pp.263–273.

[ShoJoh] J.E. Shore, R.W. Johnson *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy*, IEEE Trans Information Theory 1980;26(1):pp 26-37.

[Uff] J. Uffink, *Can the Maximum Entropy Principle be explained as a consistency Requirement?* Studies in History and Philosophy of Modern Physics 26B, p. 223 (1995).

[Zie] Gunter M. Ziegler *Lectures on Polytopes* Graduate Texts in Mathematics No.152, Springer-Verlag 1994.

Appendix A

A property of the Blaschke topology

Theorem 24 *If K is a consistent knowledge base in CL then for all $\lambda \in \mathbb{R}$, we define*

$$K^{(\lambda)} = K \cup \left\{ \sum_{i=1}^{J} v_i x_i = \lambda \right\}$$

where $v_1, \ldots v_J$ are real constants. Let $\lambda = \lambda_0$ be fixed. In the Blaschke topology, if $K^{(\lambda_0+\delta)}$ is consistent for each δ such that $0 \leq \delta \leq \delta_0$, $K^{(\lambda_0+\delta)} \to K^{(\lambda_0)}$ as $\delta \searrow 0$.

Proof Recall from the Introduction that $V^L(K)$ is a convex polytope. We express it in the form of (1.15) such that $\vec{u}^{(1)}, \vec{u}^{(2)} \ldots \vec{u}^{(m)}$ are its vertices.

Definition Let $Sumv$ be the function $Sumv : \mathbb{D}^J \to \mathbb{R}$, given by

$$Sumv(\vec{x}) = \sum_{i=1}^{J} v_i x_i \tag{A.1}$$

for all $\vec{x} \in \mathbb{D}^J$.

Now for each $t = 1, \ldots m$ we let U_t denote $Sumv(\vec{u}^{(t)})$. We assume w.l.o.g. that δ_0 is a small enough real number such that none of $U_1, U_2 \ldots U_j$ are in the interval $(\lambda_0, \lambda_0 + \delta_0)$.

Notation For all $\lambda \in \mathbb{R}$, let $Sol(\lambda)$ denote $V^L(K^{(\lambda)})$.

Lemma 166 *If $\vec{z} \in Sol(\lambda_0 + \delta)$, where δ is positive and $\delta < \delta_0$, there exists*

$\vec{x} \in Sol(\lambda_0)$ and $\vec{y} \in Sol(\lambda_0 + \delta_0)$ such that

$$\vec{z} - \vec{x} = \frac{\delta}{\delta_0}(\vec{y} - \vec{x}) \tag{A.2}$$

Remark In other words, every solution $\vec{z}$ of K such that $\sum_{i=1}^{J} v_i z_i$ is between λ_0 and $\lambda_0 + \delta_0$ is on a line segment joining a point in $Sol(\lambda_0)$ and a point in $Sol(\lambda_0 + \delta_0)$.

Proof Let $\vec{top}^{(1)}, \vec{top}^{(2)} \ldots \vec{top}^{(k)}$ enumerate the vertices of $V^L(K)$ at which the value of $Sumv$ is not less than $\lambda_0+\delta_0$ and let the other vertices (where $Sumv$ is at most λ_0) be denoted by $\vec{bot}^{(1)}, \ldots \vec{bot}^{(q)}$. Fix some $\vec{z} \in V^L(K)$ s.t. $Sumv(\vec{z}) \in (\lambda_0, \lambda_0 + \delta_0)$. Hence

$$\vec{z} = \sum_{i=1}^{k} a_i \vec{top}^{(i)} + \sum_{j=1}^{q} c_j \vec{bot}^{(j)} \tag{A.3}$$

where all a_i's and c_j's are non-negative real numbers and $\sum_{i=1}^{k} a_i + \sum_{j=1}^{q} c_j = 1$.

Claim Both $\sum_{i=1}^{k} a_i \vec{top}^{(i)}$ and $\sum_{j=1}^{q} c_j \vec{bot}^{(j)}$ must be non-zero. Thus $k > 0, q > 0$ and there exists an $a_i > 0$ and a $c_j > 0$.

Proof of claim Suppose for contradiction that $\vec{z} = \sum_{i=1}^{k} a_i \vec{top}^{(i)}$. Working out $Sumv$ gives

$$\begin{aligned} \sum_{i=1}^{J} v_i z_i = \sum_{i=1}^{J} v_i \sum_{p=1}^{k} a_p top_i^{(p)} = \sum_{p=1}^{k} a_p \sum_{i=1}^{J} v_i top_i^{(p)} &= \sum_{p=1}^{k} a_p Sumv(\vec{top}^{(p)}) \\ &\geq \sum_{p=1}^{k} a_i(\lambda_0 + \delta_0) \geq (\lambda_0 + \delta_0) \end{aligned} \tag{A.4}$$

since $\sum_{i=1}^{k} a_i = 1$, so we have found a contradiction.

The case of $\vec{z} = \sum_{j=1}^{q} c_j \vec{bot}^{(j)}$ is similar. We have proved the claim.

■

Now we can split $\vec{z}$ into its top (where $Sumv$ is at least $\lambda_0 + \delta_0$) and bottom (where $Sumv$ is at most λ_0) components.

Let

$$\vec{z}_\uparrow = \frac{\sum_{i=1}^{k} a_i \vec{top}^{(i)}}{\sum_{i=1}^{k} a_i}$$

and

$$\vec{z}_\downarrow = \frac{\sum_{j=1}^{q} c_j \vec{bot}^{(j)}}{\sum_{j=1}^{q} c_j} \tag{A.5}$$

which are well-defined by the previous claim. Since they are convex combinations of the vertices of $V^L(K)$, $\vec{z}_\uparrow$ and $\vec{z}_\downarrow$ are solutions of K. Now

$$\vec{z} = \left(\sum_{i=1}^{k} a_i\right)\vec{z}_\uparrow + \left(\sum_{j=1}^{q} c_j\right)\vec{z}_\downarrow \tag{A.6}$$

so $\vec{z}$ is on a line segment connecting two points, one where the value of $Sumv$ is $Sumv(z_\uparrow) \geq (\lambda_0 + \delta_0)$ and another where the value of $Sumv$ is $Sumv(z_\downarrow) \leq \lambda_0$. Finally, the points where $Sumv$ equals λ_0, $\lambda_0 + \delta_0$ which lie on this same line segment are

$$\vec{y} = \vec{z}_\downarrow + \frac{\lambda_0 + \delta_0 - Sumv(z_\downarrow)}{Sumv(z_\uparrow) - Sumv(z_\downarrow)}(\vec{z}_\uparrow - \vec{z}_\downarrow)$$

and

$$\vec{x} = \vec{z}_\downarrow + \frac{\lambda_0 - Sumv(z_\downarrow)}{Sumv(z_\uparrow) - Sumv(z_\downarrow)}(\vec{z}_\uparrow - \vec{z}_\downarrow) \tag{A.7}$$

respectively, which contain $\vec{z}$ between them as in (A.2). By definition, $\vec{x}$ and $\vec{y}$ are convex combinations of $\vec{z}_\uparrow$ and $\vec{z}_\downarrow$, which are solutions of K. Hence $\vec{x}$ and $\vec{y}$ are also solutions of K, by convexity of $V^L(K)$. We have proved Lemma 166.

■

Proof of Theorem 24 continued By Lemma 10, it is now sufficient for proving the theorem that

$$\Delta_{Sol(\lambda_0)\to Sol(\lambda_0+\delta)} \to 0 \text{ and } \Delta_{Sol(\lambda_0+\delta)\to Sol(\lambda_0)} \to 0 \text{ as } \delta \searrow 0 \tag{A.8}$$

For all $\vec{x}, \vec{y} \in \mathbb{D}^J$, $|\vec{x} - \vec{y}| \leq J$ so for any convex polytopes A, B in $\mathbb{D}^J$, $\Delta(A, B) \leq J$. For all $\vec{x} \in Sol(\lambda_0)$ we fix some $\vec{y} \in Sol(\lambda_0 + \delta_0)$. If we let

$$\vec{z} = \vec{x} + \frac{\delta}{\delta_0}(\vec{y} - \vec{x}) \tag{A.9}$$

then $\vec{z} \in Sol(\lambda_0 + \delta)$. That is because $V^L(K)$ is convex, so $\vec{z} \in V^L(K)$ and the linearity of $Sumv$ ensures that $Sumv(\vec{z}) = \lambda_0 + \delta$. Since $\vec{z} - \vec{x} = \frac{\delta}{\delta_0}(\vec{y} - \vec{x})$, then $|\vec{z} - \vec{x}| \leq J\frac{\delta}{\delta_0}$. As $\vec{x}$ was arbitrary,

$$\Delta_{Sol(\lambda_0)\to Sol(\lambda_0+\delta)} \leq J\frac{\delta}{\delta_0} \text{ so } \Delta_{Sol(\lambda_0)\to Sol(\lambda_0+\delta)} \to 0 \text{ as } \delta \searrow 0 \tag{A.10}$$

We now fix $\vec{z}$ such that $\vec{z} \in Sol(\lambda_0 + \delta)$. By Lemma 166, there exist

$\vec{x} \in Sol(\lambda_0), \vec{y} \in Sol(\lambda_0 + \delta_0)$ such that $\vec{z}$ is a convex combination of $\vec{x}$ and $\vec{y}$, so that

$$\vec{z} - \vec{x} = \frac{\delta}{\delta_0}(\vec{y} - \vec{x}) \tag{A.11}$$

Thus $|\vec{x} - \vec{z}| \leq J\frac{\delta}{\delta_0}$ and, since our choice of $\vec{z}$ was arbitrary, $\Delta_{Sol(\lambda_0+\delta)\to Sol(\lambda_0)} \to 0$ as $\delta \searrow 0$. Hence we have proved Theorem 24.

■

Theorem 25 *Let K be a fixed consistent knowledge base in CL and define $i_1, \ldots i_k$ s.t. $1 \leq i_1 < i_2 \ldots < i_k \leq J$. For all $\vec{\lambda} \in \mathbb{R}^k$, let $K(\vec{\lambda})$ denote $K \cup \{x_{i_1} = \lambda_1, x_{i_2} = \lambda_2, \ldots x_{i_k} = \lambda_k\}$. Then, if $\vec{\lambda}$ varies such that $K(\vec{\lambda})$ is consistent, $K(\vec{\lambda})$ is a continuous function of $\vec{\lambda}$.*

Proof We assume w.l.o.g. that $i_1 = 1, i_2 = 2, \ldots i_k = k$. Let $\vec{\lambda} = \vec{\lambda}^{(0)}$ be such that $K(\vec{\lambda}^{(0)})$ is consistent. We now show that as $\vec{\lambda} \to \vec{\lambda}^{(0)}$, $K(\vec{\lambda}) \to K(\vec{\lambda}^{(0)})$. Consider the set

$$Sector = \{\vec{x} \in V^L(K) \,|\, x_1 \geq \lambda_1^{(0)} \text{ and } x_1 - \lambda_1^{(0)} \geq |x_p - \lambda_p^{(0)}| \text{ for each } p = 2, \ldots k\} \tag{A.12}$$

which is a convex polytope containing $V^L(K(\vec{\lambda}^{(0)}))$. We now fix a $\delta_0 > 0$ such that none of the vertices $\vec{u}$ of *Sector* are such that $\lambda_1^{(0)} < u_1 < \lambda_1^{(0)} + \delta_0$.

Claim Suppose that $\vec{\lambda}$, δ, satisfy $\lambda_1 = \lambda_1^{(0)} + \delta$, $\delta \in [0, \delta_0]$ and $\lambda_1 - \lambda_1^{(0)} \geq |\lambda_p - \lambda_p^{(0)}|$ for each $p = 2, \ldots k$. If $K(\vec{\lambda})$ is consistent, then $\Delta(K(\vec{\lambda}^{(0)}), K(\vec{\lambda})) \leq J\frac{\delta}{\delta_0}$.

Proof of claim If $\vec{\lambda}$, δ satisfy the conditions of the claim then $V^L(K(\vec{\lambda})) \subseteq Sector*$ where $Sector* = \{\vec{x} \in Sector \text{ s.t. } x_1 \in [\lambda_1^{(0)}, \lambda_1^{(0)} + \delta_0]\}$.

We observe that each solution $\vec{x}$ of $V^L(K(\vec{\lambda}))$ is a convex combination of the vertices of *Sector* - some of which have the 1st co-ordinate value equal to $\lambda_1^{(0)}$ and the others have it not less than $\lambda_1^{(0)} + \delta_0$. By the definitions of *Sector* and *Sector*∗, we note that the vertices of *Sector* whose 1st co-ordinate values are equal to $\lambda_1^{(0)}$ are solutions of $K(\vec{\lambda}^{(0)})$. In a similar way to the proof of Lemma 166, we can show that every $\vec{x} \in V^L(K\vec{\lambda}^{(0)})$ is a convex combination of a solution of $K(\vec{\lambda}^{(0)})$ and some solution of a $K(\vec{b})$ for which $b_1 = \lambda_1^{(0)} + \delta_0$.

These conditions uniquely specify $\vec{b}$ independently of $\vec{x}$. Indeed $\vec{b}$ is the only vector on the line connecting $\vec{\lambda}^{(0)}$ and $\vec{\lambda}$ at which the 1st co-ordinate has value $\lambda_1^{(0)} + \delta_0$. Hence every solution of $K(\vec{\lambda})$ is a convex combination of a solution of $K(\vec{\lambda}^{(0)})$ and a solution of $K(\vec{b})$, where $\vec{b}$ is independent of the choice of $\vec{x}$.

Proof of Theorem 25 Now the argument proceeds as it did for Theorem 24 to deduce that

$$\Delta_{V^L(K(\vec{\lambda}^{(0)}))\rightarrow V^L(K(\vec{\lambda}))} \text{ and } \Delta_{V^L(K(\vec{\lambda}))\rightarrow V^L(K(\vec{\lambda}^{(0)}))} \tag{A.13}$$

are bounded by $J\frac{\delta}{\delta_0}$. This proves that the function $K(\lambda_1, \ldots \lambda_k)$ is continuous at $\vec{\lambda}^{(0)}$ in the region of those $\vec{\lambda}$ whose first co-ordinate value is greater than $\lambda_1^{(0)}$ and such that no other λ_p is further from $\lambda_p^{(0)}$. That region is, in terms of the argument above, similar to the other $2k-1$ regions making up the set of all $\vec{\lambda}$ that make $K(\vec{\lambda})$ consistent: namely pick a co-ordinate value of $\vec{\lambda}$ most different to its value in $\vec{\lambda}^{(0)}$ and state whether that difference is positive or negative.

Since continuity of $K(\lambda)$ at $\vec{\lambda}^{(0)}$ is true in a finite number of regions covering the neighbourhood around $\vec{\lambda}^{(0)}$, the function is continuous at $\vec{\lambda}^{(0)}$ in the complete neighbourhood of that point. Thus we have proved the theorem.

■