Abstract

UNIVERSITY OF MANCHESTER

ABSTRACT OF DISSERTATION submitted by **Jonathan Lawry** for the Degree of Ph.D. and entitled **Inexact Reasoning**

Month and Year of Submission: May 1994

The problem of modelling inductive inference on knowledge bases expressing an experts degree of belief in certain statements from a language is introduced. The type of knowledge base considered is then restricted to those consisting of a set of linear constraints on a subjective probabilistic belief function over a finite language of the propositional calculus with connectives $\wedge, \vee, \neg$.

The notion of a second order prior distribution on the set of probabilistic belief functions on a language is then defined together with a method of updating from knowledge bases to give a possible inference proceedure. The problem of selecting a particular prior distribution for a language with n propositional variables is restricted to that of picking a smooth prior density on a $2^n - 1$ dimensional

simplex in $\mathbb{R}^{2^n}$. The relationship between a number of epistemological principles and properties of hierarchies of such priors is demonstrated. In particular an epistemological characterisation of certain hierarchies of symmetric Dirichlet priors upto a parameter is given.

The above model of inductive inference is related to that of an inference process by considering the expected probabilistic belief function relative to a second order prior distribution and a knowledge base .This corresponds to the center of mass of a convex subset of the afore mentioned simplex.

Chapter 1

Motivations and Formulation

1.1 Introduction

Throughout this dissertation the notion of belief is modelled in a quantitative way. Futhermore it is assumed that a weight value between 0 and 1 can be allocated to any sentence of a given language to represent an agents level of belief in the assertion that the sentence is true. This assumption is not, in fact, universally accepted as a sound basis for modelling human reasoning, it having been argued that generally human beings are not able to make such numerical allocations in a meaningful or even consistent way. The claim has even been made that beliefs in statements from any reasonably expressive language can at best be partially ordered. However, although we concede that it is often difficult for people to give precise values to their beliefs we feel that this is due more to ignorance regarding the meaning of the statements they are asked to consider and the relationships between them than the nature of human reasoning itself. Certainly if individuals are questioned on a topic in which they have strong expertise it seems to be possible for them to convert linguistic statements into ones concerning belief values.

To formalise this notion of belief we restrict ourselves to the propositional calculus with connectives $\wedge$, $\vee$, $\neg$. Let L be a countably infinite language of the propositional calculus with the above connectives and some enumeration of the propositional variables. Then let $L^{(n)} = \{p_1, \ldots, p_n\}$ these being the first n propositional variables in the enumeration of L. The sentences of $L^{(n)}$ are denoted by $SL^{(n)}$. We define a belief function and conditional belief function on $SL^{(n)}$ as follows. For θ , $\phi \in SL^{(n)}$ let $Bel(\theta)$ be the belief value in θ and $Bel(\theta|\phi)$ the conditional belief value in θ given ϕ so that;

$$Bel() : SL^{(n)} \to [0, 1]$$

$$Bel(|) : SL^{(n)} \times SL^{(n)} \to [0, 1]$$

Now a number of different sets of principles have been proposed that Bel should satisfy if it is to model natural inexact reasoning. These classify different types of belief function but for this thesis we shall be concerned only with probabilistic belief functions.

1.2 Belief as Probability

Let $\overline{SL^{(n)}}$ be the Lindenbaum algebra of equivalence classes under the equivalence relation of logical equivalence generated by $SL^{(n)}$ (ie for $\theta, \varphi \in SL^{(n)}$ the algebra consists of equivalence classes $\overline{\theta}$ such that $\varphi \in \overline{\theta}$ iff $\models \varphi \leftrightarrow \theta$). The largest and smallest elements of this algebra are denoted by $\top$ and $\bot$ respectively. Now by the disjunctive normal form theorem (see [17] [7]) every sentence of $L^{(n)}$ is equivalent to a disjunction of sentences of the form $\bigwedge_i^n p_i^{\epsilon_i}$ where $\epsilon_i \in \{0,1\}$ and p^1 represents p, p^0 represents $\neg p$. The equivalence classes containing these sentences form the atoms of $\overline{SL^{(n)}}$ relative to the natural ordering on boolean algebras and hence we refer to them as atoms. For a language of size n there are 2^n atoms which we enumerate by

$$\alpha_j^{(n)} = \bigwedge_{i=1}^{n} p_i^{\epsilon_i^j}$$

where

$$j = 2^n - \sum_{i=1}^{n} \epsilon_i^j 2^{n-i}$$

Clearly the $\bar{\alpha}_i^{(n)}$ are disjoint classes and $\bigvee_i^{2^n} \bar{\alpha}_i^{(n)} \ = \top$.

Definition 1 $\overline{Bel} : \overline{SL^{(n)}} \to [0,1]$ *is a probability measure on* $\overline{SL^{(n)}}$ *if the following are satisfied:*

(i) $\overline{Bel}(\top) = 1$

(ii) $\overline{Bel}(\overline{\theta} \vee \overline{\varphi}) = \overline{Bel}(\overline{\theta}) + \overline{Bel}(\overline{\varphi})$ *if* $\overline{\theta} \wedge \overline{\varphi} = \bot$

The corresponding conditional probability measure is then defined as follows;

Definition 2 $\forall \overline{\theta}, \overline{\varphi} \in \overline{SL^{(n)}}$ *such that* $\overline{Bel}(\overline{\varphi}) > 0$

$$\overline{Bel}(\overline{\theta}|\overline{\varphi}) = \frac{\overline{Bel}(\overline{\theta} \wedge \overline{\varphi})}{\overline{Bel}(\overline{\varphi})}$$

Now any probability measure $\overline{Bel}$ on $\overline{SL^{(n)}}$ uniquely defines a belief function and conditional belief function on $SL^{(n)}$ such that

$$\forall \theta \in SL^{(n)} \; Bel(\theta) = \overline{Bel}(\overline{\theta})$$

$$\forall \theta, \varphi \in SL^{(n)} \; Bel(\varphi) > 0 \; then \; Bel(\theta|\varphi) = \frac{Bel(\theta \wedge \varphi)}{Bel(\varphi)}$$

It follows then from the definition of $\overline{Bel}$ that any Bel defined in this way will satisfy

(i') If $\models \theta$ then $Bel(\theta) = 1$

(ii') if $\models \neg(\theta \wedge \varphi)$ then $Bel(\theta \vee \varphi) = Bel(\theta) + Bel(\varphi)$

Also, since (i') and (ii') together imply

(iii') if $\models \theta \longleftrightarrow \varphi$ then $Bel(\theta) = Bel(\varphi)$

we have that any belief function defined according to (i') and (ii') uniquely defines a probability measure on $\overline{SL^{(n)}}$. Thus we define a probabilistic belief function to be any belief function satisfying (i') and (ii') and for $\theta \in SL^{(n)}$ $Bel(\theta)$ is interpreted as the subjective probability of θ.

1.3 Random Propositional variables and Random Formulae

An alternative formulation of probabilistic belief is obtained in the following manner.

Definition 3 *$\underline{p}$ is a Random Propositional Variable if $\underline{p}$ is a random variable with possible outcomes 0 and 1 such that $\underline{p}$ is interpreted as being the truth value of p a propositional variable of $L^{(n)}$ for some propositional language as above.*

$\underline{\theta}$ is a Random Formula if $\underline{\theta}$ is a random variable with possible outcomes 0 and 1 where $\underline{\theta}$ is interpreted as being the truth value of $\theta \in SL^{(n)}$.

Clearly for $\theta \in SL^{(n)}$ $\underline{\theta} = F_\theta(\underline{p_1}, \ldots, \underline{p_n})$ where $F_\theta : \{0,1\}^n \to \{0,1\}$ is defined according to the standard truth functional interpretations of the connectives $\wedge, \vee, \neg$ in propositional logic.

Given a probability distribution over all random formula $\underline{\theta}$ such that $\theta \in SL^{(n)}$ $E(\underline{\theta}) =$ *probability* $\underline{\theta} = 1$. Therefore, if $Bel(\theta) = E(\underline{\theta})$ then Bel satisfies (i') and (ii'). In the sequel we adopt this notation although for the sake of elegance we drop the bar and simply write $E(\theta)$.

1.4 Knowledge Bases

Consider the following situation where a doctor with strong expertise in the diagnosis of a certain disease D is asked to provide a sequence of statements expressing his general knowledge in this area, or at least a fragment of it. Then he might well give the following replies:

Symptom S strongly suggests D

Patients of blood group B rarely develop D

Patients of blood group B with symptom S are much more likely than not to have disease D

A small proportion of patients are of blood group B

It seems apparent then that the doctor is revealing a set of rules and background information which he considers fundamental to his reasoning process. By the latter we mean the process by which he is able to make judgements about his belief in statements concerning S, B and D which do not appear above. We shall discuss later a notion of inductive inference that might be taking place here but initially let us consider how we might formalise the above statements.

Now although the meanings of natural language terms such as 'strongly suggests' and 'rarely develop' are ambiguous they do seem to refer to degrees of belief or conditional belief. Therefore, in the context of probabilistic belief functions it seems natural to attempt to encode these statements as constraints on some probability function E. For example, one interpretation might be;

$$\begin{aligned} & E(D|S) && > 0.75 \\ & E(D|B) && < 0.25 \\ & E(D|B \wedge S) && > 0.7 \\ 0.02 < \ & E(B) && < 0.05 \end{aligned}$$

Alternatively, you could argue that a coding in terms of equality constraints is more appropriate here. A possibility is;

$$\begin{aligned} E(D \wedge S) &= 0.8E(S) \\ E(D \wedge B) &= 0.1 \\ E(D \wedge B \wedge S) &= 0.85 \\ E(B) &= 0.03 \end{aligned}$$

In the following both interpretations for these types of linguistic statements are considered. That is we propose models of inference on sets of equality constraints and sets of strict inequality constraints.

In a practical situation where some form of automated reasoning procedure (eg expert system) was being developed to aid diagnosis of disease D it might be hoped that the doctor himself would be able to help to express the knowledge base in this form.

Knowledge bases for expert systems often consist of a set of statements such as those above and hence a pertinent question would seem to be what is a reasonable inference procedure based on such knowledge bases coded as constraints on a probabilistic belief function. Much work has been done on this by Paris and Vencovska (see [19] and [22]) but we shall adopt a somewhat different approach.

Throughout the sequel the expression interval of $\mathbb{R}$ will be used to mean a set of the form $[c, d]$, (c, d) , $[c, d)$, $(c, d]$ or $\{d\}$ where $c, d \in \mathbb{R}$ $c < d$. A strict interval of $\mathbb{R}$ will refer to an interval of $\mathbb{R}$ which is not a singleton set. In addition, an open interval of $\mathbb{R}$ will mean a set (c, d) where $c < d$. An interval of $[0, 1]$ will naturally refer to an interval of $\mathbb{R}$ where $0 \leq c < d \leq 1$ and the expressions strict interval of $[0, 1]$ and open interval of $[0, 1]$ will be taken to have the obvious meanings.

Further, if $I_1 \equiv (c, d)$, $I_2 \equiv [c, d]$, $I_3 \equiv (c, d]$ and $I_4 \equiv [c, d)$ for $c < d$ $c, d \in \mathbb{R}$ then for $x \in \mathbb{R}^{\geq 0}$ xI_1, xI_2, xI_3 and xI_4 will refer respectively to the intervals (xc, xd), $[xc, xd]$, $(xc, xd]$ and $[xc, xd)$ and $0I_j$ will refer to $\{0\}$ for $j = 1, \ldots, 4$.

The state of an agent's belief on the sentences of a given language is assumed to be expressible in terms of sets of linear interval constraints on their probabilistic belief function and of interval contraints on their conditional belief function. More formally:

Definition 4 *A Probability State of $L^{(n)}$ is a set of constraints of the form*

$$\sum_{j=1}^{m_i} a_{i,j} E(\theta_i) \in I_i \ for \ i = 1, \ldots, k$$
$$E(\theta_i \wedge \varphi_i) \in E(\varphi_i) I_i \ for \ i = k+1, \ldots, l$$

where $k, l \in \mathbb{N}$ $k \geq l$, I_i is an interval of $\mathbb{R}$, $a_{i,j} \in \mathbb{R}$ for $j = 1, \ldots, m_i$, $i = 1, \ldots, k$, I_i is an interval of $[0,1]$ for $i = k+1, \ldots, l$, $\theta_i, \varphi_i \in SL^{(n)}$ and E is a probabilistic belief function on $SL^{(n)}$

In particular we shall consider inference on two types of probability state.

Definition 5 *An Open Interval Probability State of $L^{(n)}$ is a set of constraints of the following form*

$$c_i \ < \ \sum_{j=1}^{m_i} a_{i,j} E(\theta_i) \ < \ d_i \ for \ i = 1, \ldots, k$$
$$c_i \ < \ E(\theta_i | \varphi_i) \ < \ d_i \ and \ E(\varphi_i) \neq 0 \ for \ i = k+1, \ldots, l$$

where $k, l \in \mathbb{N}$ $k \geq l$, $c_i, d_i, a_{i,j} \in \mathbb{R}$ $c_i < d_i$ for $j = 1, \ldots, m_i$ $i = 1, \ldots, k$ and $0 \leq c_i < d_i \leq 1$ for $i = k+1, \ldots, l$ and E is a probabilistic belief function on $SL^{(n)}$.

Definition 6 *An Equality Probability State of $SL^{(n)}$ is a set of constraints of the form*

$$\sum_{j=1}^{m_i} a_{i,j} E(\theta_i) = b_i \ for \ i = 1, \ldots, l$$

where $l \in \mathbb{N}$ $a_{i,j}, b_i \in \mathbb{R}$ for $j = 1, \ldots, m_i$ and $i = 1, \ldots, l$ and E is a probabilistic belief function on $SL^{(n)}$.

Definition 7 *A Probability State C' is the atomic form of C if it is the result of replacing every occurrence of θ in C with $\bigvee_{i\in S_\theta} \alpha_i^{(n)}$ where $S_\theta = \{i \in \{1,\dots,2^n\} | \alpha_i^{(n)} \models \theta\}$*

Thus we say that C is in atomic form if the atomic form of C is C.

A probability function E on $SL^{(n)}$ is said to be consistent with a probability state C if E satisfies every constraint in C. A probability state C is consistent if there exists a probability function on $SL^{(n)}$ consistent with C

Definition 8

$$KB(L^{(n)}) = \{C|\ C\ is\ a\ consistent\ probability\ state\ of\ L^{(n)}\}$$

$$IKB(L^{(n)}) = \{C|\ C\ is\ a\ consistent\ open\ interval\ probability\ state\ of\ L^{(n)}\}$$

$$EKB(L^{(n)}) = \{C|\ C\ is\ a\ consistent\ equality\ probability\ state\ of\ L^{(n)}\}$$

A **Knowledge Base** is a probability state upon which some inference is conditional. The terms Open Interval Knowledge base and Equality Knowledge Base are then taken to have the obvious meanings. In the sequel the capital letters C and K will be used to denote probability states and knowledge bases respectively.

A possible justification for interpreting knowledge as open interval probability states is as follows. If it is supposed that an agent obtains his general knowledge from experience then it is more natural to interpret this knowledge in terms of strict inequality constaints than equality constraints since most experience is only likely to enable the agent to put some bounds on his beliefs.

In addition, if an agents knowledge contains a constraint on conditional belief of the form

$$E(\theta|\varphi) \in (c, d)\ 0 \leq c < d \leq 1\ \theta, \varphi \in SL^{(n)}$$

then it seems natural to assume, in this context, that he must also know that $E(\varphi) > 0$. This is because if the agents information comes from experience rather than being theoretical then to know anything regarding θ in situations when the event $\underline{\varphi} = 1$ has occurred then he must have experienced such situations. Clearly, however, this argument is not applicable in the case where φ is a tautology since for any $\theta \in SL^{(n)}$ where it is known that $E(\theta) \in (c, d)$ it can be deduced that $E(\theta|\varphi) \in (c, d)$ without having any empirical knowledge relating specifically to φ. However, it would seem somewhat unnatural to explicitly state the latter conditional probability in the knowledge base. Hence, in this model of inequality knowledge bases it is assumed that all constraints result from empirical evidence only rather than being deduced from other constraints and *a priori* knowledge relating to logical structure.

As suggested by the above comments we consider equality knowledge bases to be more appropriate when the knowledge is theoretical in nature. However, Paris has argued that it is reasonable to generally interpret knowledge bases for Inexact Reasoning as equality probability states. See Paris [22] for a general discussion.

1.5 A Geometric Representation of Probability States

For any non empty propositional language there are *a priori* uncountably many probabilistic belief functions that can be defined. Even when a knowledge base is specified it is generally the case that it does not uniquely characterise one or even countably many belief functions.

Definition 9

$$PF(L^{(n)}) = \{E|\ E\ is\ a\ probability\ function\ on\ SL^{(n)}\}$$

$$V^{(n)} = \{\vec{x} \in [0,1]^{2^n} | \sum_{i=1}^{2^n} x_i = 1\} \subset [0,1]^{2^n}$$

A point $\vec{x} \in V^{(n)}$ is said to satisfy a probability state C if $\vec{x}$ satisfies every constraint formed by substituting $\sum_{i \in S_\theta} x_i$ for every occurrence of $E(\theta)$, for all θ in $SL^{(n)}$, in the constraints in C.

Definition 10

$$PF(L^{(n)})(C) = \{E \in PF(L^{(n)})|\ E\ is\ consistent\ with\ C\}$$

$$V^{(n)}(C) = \{\vec{x} \in V^{(n)}|\ \vec{x}\ satisfies\ C\}$$

By the disjunctive normal form theorem of propositional logic there is a natural correspondence between points in $V^{(n)}$ and probability functions in $PF(L^{(n)})$ in that any point $\vec{x} \in V^{(n)}$ uniquely determines a probability function E on $SL^{(n)}$ for which $E(\alpha_i^{(n)}) = x_i$ and vica versa. In the same way, for any probability state C there is a correspondence between points $V^{(n)}(C)$ and probability functions in $PL(L^{(n)})(C)$.

Definition 11 *Two probability states C and C' are said to be equivalent if $V^{(n)}(C) = V^{(n)}(C')$*

Clearly any C is equivalent to the atomic form of C. Therefore, in the sequel we shall generally consider only probability states in atomic form.

Proposition 12 *If C is a Probability State then $V^{(n)}(C)$ is convex.*

Further, for an equality probability state C if $\vec{x}, \vec{y} \in V^{(n)}(C)$ and $\lambda\vec{x} + (1-\lambda)\vec{y} \in V^{(n)}$ then $\lambda\vec{x} + (1-\lambda)\vec{y} \in V^{(n)}(C)$ for $\lambda \in \mathbb{R}$

Proof

For Equality probability states the result is proved in Paris [22]

We shall prove the result only for open interval probability states since the structure of the proof is identical for sets of other interval constraints

Suppose C is an open interval probability state then C is equivalent to a set of constraints

$$c_i < \sum_{j=1}^{2^n} a_{i,j} E(\alpha_j^{(n)}) < d_i \quad i = 1, \ldots, k$$

$$c_i < \frac{\sum_{j \in S_{\theta_i \wedge \varphi_i}} E(\alpha_j^{(n)})}{\sum_{j \in S_{\varphi_i}} E(\alpha_j^{(n)})} < d_i$$

and $\sum_{j \in S_{\varphi_i}} E(\alpha_j^{(n)}) > 0$ for $i = k+1, \ldots, l$

where $k, l \in \mathbb{N}$ $k \leq l$ $c_i, d_i, a_{i,j} \in \mathbb{R}$ for $i = 1, \ldots, k$ $j = 1, \ldots, 2^n$ and $0 \leq c_i < d_i \leq 1$ for $i = k+1, \ldots, l$ and E is a probability function. Hence,

$\forall \vec{x} \in V^{(n)}\ \vec{x} \in V^{(n)}(C)$ iff $\vec{x}$ satisfies the constraints

$$c_i < \sum_{j=1}^{2^n} a_{i,j} x_j < d_i \quad i = 1, \ldots, k$$

$$c_i < \frac{\sum_{j \in S_{\theta_i \wedge \varphi_i}} x_j}{\sum_{j \in S_{\varphi_i}} x_j} < d_i \quad i = k+1, \ldots, l$$

Now suppose $\vec{x}, \vec{y} \in V^{(n)}(C)$ then consider $\lambda \vec{x} + (1-\lambda)\vec{y}$ for $\lambda \in [0,1]$. For $i = 1, \ldots, k$

$$\sum_{j=1}^{2^n} a_{i,j}(\lambda x_j + (1-\lambda) y_j) = \lambda \sum_{j=1}^{2^n} a_{i,j} x_j + (1-\lambda) \sum_{j=1}^{2^n} a_{i,j} y_j$$

and

$$c_i = \lambda c_i + (1-\lambda) c_i < \lambda \sum_{j=1}^{2^n} a_{i,j} x_j + (1-\lambda) \sum_{j=1}^{2^n} a_{i,j} y_j < \lambda d_i + (1-\lambda) d_i = d_i$$

Also, for $i = k+1, \ldots, l$ clearly $\sum_{j \in \varphi_i} \lambda x_j + (1-\lambda) y_j > 0$ and

$$\sum_{j \in S_{\varphi_i \wedge \theta_i}} \lambda x_j + (1-\lambda) y_j = \lambda \sum_{j \in S_{\varphi_i \wedge \theta_i}} x_j + (1-\lambda) \sum_{j \in S_{\varphi_i \wedge \theta_i}} y_j < \lambda d_i \sum_{j \in S_{\varphi_i}} x_j + (1-\lambda) d_i \sum_{j \in S_{\varphi_i}} y_j$$

$$= d_i \sum_{j \in \varphi_i} \lambda x_j + (1-\lambda) y_j$$

and similarly

$$\sum_{j \in S_{\varphi_i \wedge \theta_i}} \lambda x_j + (1-\lambda) y_j > c_i \sum_{j \in S_{\varphi_i}} \lambda x_j + (1-\lambda) y_j$$

$\square$

Definition 13 *The dimension of a convex set $S \subseteq \mathbb{R}^n$ for $n \in \mathbb{N}$ is the largest $r \in \mathbb{N}$ such that S contains $r+1$ points $\vec{x}_{(1)}, \ldots, \vec{x}_{(r+1)}$ for which $\vec{x}_{(2)} - \vec{x}_{(1)}, \ldots, \vec{x}_{(r+1)} - \vec{x}_{(1)}$ are linearly independent.*

Definition 14 *(i) $D^{(n)} = \{\vec{x} \in \mathbb{R}^{2^n} | \ \sum_{i=1}^{2^n} x_i = 1\}$*

(ii) If C is probability state then $D^{(n)}(C) = \{\vec{x} \in D^{(n)} \mid \vec{x} \ satisfies \ C\}$

(iii) $N(\vec{x} : \epsilon) = D^{(n)} \cap \{\vec{y} \in \mathbb{R}^{2^n} \mid |\vec{x} - \vec{y}| < \epsilon\}$ and $S \subseteq D^{(n)}$ is open in $D^{(n)}$ if $\forall \vec{x} \in S \ \exists \epsilon > 0$ such that $N(\vec{x} : \epsilon) \subseteq S$

(iv) $Int(V^{(n)}) = \{\vec{x} \ \in V^{(n)} \mid x_i > 0 \ , \ i = 1, \ldots 2^n\}$

(v) $B(V^{(n)}) = V^{(n)} - Int(V^{(n)})$

Proposition 15 *If C is an Open Interval Probability State then $V^{(n)}(C) \cap Int(V^{(n)})$ is open in $D^{(n)}$ and further $V^{(n)}(C) \cap Int(V^{(n)}) = \emptyset \ \ iff \ \ V^{(n)}(C) = \emptyset$*

Lemma 16 *If $C = \{ \ c < \sum_{j=1}^{2^n} a_j E(\alpha_j^{(n)}) < d\}$ for $c, d \in \mathbb{R}, c < d$ $a_j \in \mathbb{R}$ then $D^{(n)}(C)$ is open in $D^{(n)}$*

Proof

Suppose $D^{(n)}(C) \neq \emptyset$ then for $\vec{x} \in D^{(n)}(C)$ let $\sum_{j=1}^{2^n} a_j x_j = b$ such that $b \in (c, d)$

Now suppose $\vec{y} \in N(\vec{x} : \epsilon)$ for some $\epsilon > 0$ then $x_i - y_i = \epsilon_i$ for $i = 1, \ldots, 2^n$ where $|\epsilon_i| < \epsilon$ then

$$\sum_{j=1}^{2^n} a_j y_j = \sum_{j=1}^{2^n} a_j x_j - \sum_{j=1}^{2^n} a_j \epsilon_j = b - \sum_{j=1}^{2^n} a_j \epsilon_j$$

now $|\sum_{j=1}^{2^n} a_j \epsilon_j| < \epsilon |\sum_{j=1}^{2^n} a_j|$

$$\therefore \forall \delta > 0 \ \exists \epsilon > 0 \ \ \textit{such that} \ \sum_{j=1}^{2^n} a_j y_j \in (b - \delta, b + \delta)$$

Furthermore $\exists \delta > 0$ such that $(b - \delta, b + \delta) \subseteq (c, d)$. $\therefore \exists \epsilon' > 0$ such that $\forall \epsilon < \epsilon'$ $N(\vec{x} : \epsilon) \subseteq D^{(n)}(C)$ $\Box$

Lemma 17 *If $C = \{c < E(\theta|\varphi) < d \ , E(\varphi) > 0\}$ for $c, d \in [0, 1], c < d$ then $D^{(n)}(C)$ is open in $D^{(n)}$*

Proof

Suppose $D^{(n)}(C) \neq \emptyset$ then let $\vec{x} \in D^{(n)}(C)$ and $X = \sum_{j \in S_{\theta \wedge \varphi}} x_j \quad T = \sum_{j \in S_\varphi} x_j$ then since $T > 0$ $\frac{X}{T} \in (c, d)$

Now since (c, d) is an open interval and $\exists \delta > 0$ such that $(\frac{X}{T} - \delta, \frac{X}{T} + \delta) \subseteq (c, d)$

Let $\vec{y} \in N(\vec{x} : \epsilon)$ then $x_i - y_i = \epsilon_i$ where $|\epsilon_i| < \epsilon$ for $i = 1, \ldots, 2^n$

Let $X' = \sum_{j \in S_{\theta \wedge \varphi}} y_j$ and $T' = \sum_{j \in S_\varphi} y_j$ then $X' = X - \sum_{j \in S_{\theta \wedge \varphi}} \epsilon_j$ and $T' = T - \sum_{j \in S_\varphi} \epsilon_j$ $\therefore$ for ϵ sufficiently small

$$|\frac{X}{T} - \frac{X'}{T'}| = |\frac{X}{T} - \frac{X - \sum_{j \in S_{\theta \wedge \varphi}} \epsilon_j}{T - \sum_{j \in S_\varphi} \epsilon_j}| = |\frac{X(T - \sum_{j \in S_\varphi} \epsilon_j) - T(X - \sum_{j \in S_{\theta \wedge \varphi}} \epsilon_j)}{T(T - \sum_{j \in S_\varphi} \epsilon_j)}|$$

$$< \frac{|X(\sum_{j \in S_\varphi} \epsilon_j)| + |T(\sum_{j \in S_{\theta \wedge \varphi}} \epsilon_j))|}{T^3} < \frac{(X + T)|S_\varphi|\epsilon}{T^3}$$

$$\therefore \exists \epsilon' > 0 \ \textit{such that} \ \forall \epsilon < \epsilon' \ |\frac{X}{T} - \frac{X'}{T'}| < \delta \ \Rightarrow \ \frac{X'}{T'} \in (c, d) \Rightarrow \vec{y} \in D^{(n)}(C)$$

$\square$

Proof of Proposition 15

Now if C' is an Open Interval probability state then C' is equivalent to some C such that $\forall \vec{x} \in V^{(n)}$ $\vec{x} \in V^{(n)}(C)$ iff $\vec{x}$ satisfies the constraints

$$c_i < \sum_{j=1}^{2^n} a_j x_j < d_i \quad i = 1, \dots k$$

$$c_i < \frac{\sum_{j \in S_{\theta \wedge \varphi}} x_i}{\sum_{j \in S_\varphi} x_i} < d_i \ , \quad \sum_{j \in S_\varphi} x_i > 0 \quad i = k+1, \dots, l$$

where $k, l \in \mathbb{N}$ and $k \leq l$

Clearly, then by the previous lemmas $D^{(n)}(C') \cap Int(V^{(n)}) = D^{(n)}(C) \cap Int(V^{(n)})$ is open in $D^{(n)}$

Suppose $D^{(n)}(C) \cap Int(V^{(n)}) = \emptyset$ and $V^{(n)}(C) \neq \emptyset$ then $\exists \vec{y} \in V^{(n)}$ such that $\vec{y} \in D^{(n)}(C)$

Now $D^{(n)}(C)$ is open by the above lemmas $\therefore \exists \epsilon$ such that $N(\vec{y} : \epsilon) \subseteq D^{(n)}(C)$

But $\forall \epsilon > 0$ $\vec{y} \in V^{(n)}$ $N(\vec{y} : \epsilon) \cap Int(V^{(n)}) \neq \emptyset$ because $\vec{y} \in B(V^{(n)})$ and this implies by the convexity of $V^{(n)}$ that $\lambda \vec{y} + (1 - \lambda) < \frac{1}{2^n}, \dots, \frac{1}{2^n} > \in Int(V^{(n)})$ for all $\lambda \in (0, 1)$

Also $\forall \epsilon$ $\exists \lambda \in (0, 1)$ such that $\lambda \vec{y} + (1 - \lambda) < \frac{1}{2^n}, \dots, \frac{1}{2^n} > \in N(\vec{y} : \epsilon)$

This is a contradiction

$\therefore V^{(n)}(C) = \emptyset$ $\square$

Corollary 18 *If C is a consistent Open Interval Probability State then* $dim(V^{(n)}(C)) = 2^n - 1$

Proof

Since C is a consistent Open Interval Probability state we have by proposition 15 that $D^{(n)}(C) \cap Int(V^{(n)})$ is open in $D^{(n)}$ and $D^{(n)}(C) \cap Int(V^{(n)}) \neq \emptyset$

Therefore, let $\vec{x} \in V^{(n)}(C) \cap Int(V^{(n)})$. Then $\exists \epsilon > 0$ such that $N(\vec{x} : \epsilon) \subseteq V^{(n)}(C) \cap Int(V^{(n)})$

Now $N(\vec{x} : \epsilon)$ is convex and $dim(N(\vec{x} : \epsilon)) = 2^n - 1 \Rightarrow \; dim(V^{(n)}(C)) \geq 2^n - 1$

$\therefore \; dimV^{(n)}(C) = 2^n - 1 \; \Box$

1.6 Inference Processes

I return now to the problem of how to model inference from the type of knowledge we have been considering. Now it would be required of any effective reasoning procedure to extend in some way the information contained in the knowledge base. In the context of probabilistic belief this could be interpreted to mean restricting the set of belief functions consistent with the knowledge by making *a priori* assumptions. In fact we insist that an inference process selects a unique probability function consistent with the knowledge base.

Definition 19 *(i) An Open Interval Inference Process relative to a language $L^{(n)}$ is a function*

$$N(L^{(n)}) : IKB(L^{(n)}) \; \rightarrow \; PF(L^{(n)})$$

such that $N(L^{(n)})(K) \in PF(L^{(n)})(K)$

(ii) An Equality Inference Process relative to a language $L^{(n)}$ is a function

$$N(L^{(n)}) : EKB(L^{(n)}) \;\rightarrow\; PF(L^{(n)})$$

such that $N(L^{(n)})(K) \in PF(L^{(n)})(K)$

For all $\theta \in SL^{(n)}$ we write $N_K(L^{(n)})(\theta)$ for the belief value of θ given by $N(L^{(n)})(K)$

The term Inference Process will be used to refer to both Equality and Open Interval Inference Processes.

Let $dom_N(L^{(n)})$ be the domain of the inference process $N(L^{(n)})$ so that either $dom_N(L^{(n)}) = EKB(L^{(n)})$ or $dom_N(L^{(n)}) = IKB(L^{(n)})$

Definition 20 *A hierarchy of inference processes is a sequence of equality inference processes or a sequence of open interval inference processes $\{N(L^{(n)})\}$ where $n \in \mathbb{N}\ ,\ n > 0$*

A number of different principles on inference process have been proposed by Paris and Vencovska (see [19] and [22]) that are intended to be characteristic of the reasoning of some ideally rational agent. Below we give details of those which will be of relevance.

Equivalence Principle

An inference process $N(L^{(n)})$ satisfies the equivalence principle if for equivalent $K_1, K_2 \in dom_N(L^{(n)})$ we have $\forall \theta \in SL^{(n)}$

$$N_{K_1}(L^{(n)})(\theta) = N_{K_2}(L^{(n)})(\theta)$$

In other words, the way in which knowledge is expressed within a language should not affect inferences on that knowledge. In the sequel, all inference processes considered will have this property .

Invariance Under Renaming

Let σ be a permutation of $\{1, \ldots, 2^n\}$. For $K \in dom_N(L^{(n)})$ such K is in atomic form let K_σ be the result of replacing each occurrence of $\alpha_i^{(n)}$ for $i = 1, \ldots, 2^n$ in the constraints of K with $\alpha_{\sigma(i)}^{(n)}$. Then $N(L^{(n)})$ is invariant under renaming if $\forall \theta \in SL^{(n)}$

$$N_K(L^{(n)})(\theta) = N_{K_\sigma}(L^{(n)})(\theta_\sigma)$$

The justification for this principle is that the atoms of $SL^{(n)}$ all share the same status of being simply possible world and so the particular ordering of the atoms which we choose should not be significant.

Language Invariance

A hierarchy of inference processes $\{N(L^{(n)})\}$ is language invariant if $\forall n > 0$, $K \in dom_N(L^{(n)})$ and $\forall \theta \in SL^{(n)}$

$$N_K(L^{(n)})(\theta) = N_K(L^{(n+1)})(\theta)$$

For inferences satisfying this principle adding a new propositional variable to the language about which nothing is known does not affect the belief value inferred for sentences not containing this propositional variable.

Irrelevant Information

A hierarchy of inference processes $\{N(L^{(n)})\}$ satisfies irrelevant information if $\forall n, m > 0$ $K_1 \in dom_N(L^{(n)})$ $K_2 \in dom_N(L^{(n+m)} - L^{(n)})$ $\theta \in SL^{(n)}$

$$N_{K_1+K_2}(L^{(n+m)})(\theta) = N_{K_1}(L^{(n)})(\theta)$$

The justification for this principle is that the knowledge provided by K_2 is irrelevant to K_1 and θ since it concerns a completely separate set of propositional variables.

We propose the following additional principle for equality inference processes

Zero Probability Invariance (ZP Invariance)

A hierarchy of equality inference processes $\{N(L^{(n)})\}$ is Zero Probability Invariant if $\forall n > 0$ $\forall K \in EKB(L^{(n)})$ $\forall \theta \in SL^{(n)}$

$$N_K(L^{(n)})(\theta) = N_{K+\{E(p_{n+1})=0\}}(L^{(n+1)})(\theta)$$

Less formally this principle states that adding a new propositional variable to the language together with only the knowledge that it has zero probability should not affect the belief value inferred for sentences not containing that propositional variable

Note that for all hierarchies of inference processes Irrelevant Information implies Language Invariance and for hierarchies of equality inference processes Irrelevant Information implies ZP invariance.

1.7 Second Order Probability

If a human expert is asked to give his or her belief in some statement relative to a fixed knowledge base on different occasions he or she is unlikely to give exactly the same responses. This might be interpreted as being the result of some error arising in the inference procedure but we feel that it is more likely to be a manifestation of uncertainties inherent in the reasoning process itself. We propose then that the answers given by the expert were being generated according to some underlying second order distribution on beliefs consistent with the knowledge base. That is to say that the expert's beliefs relative to a knowledge base are generally not calculated according to some inference process as defined in the previous section. Instead they are generated according to a distribution conditional on the knowledge base which is determined by some updating procedure from an initial prior. In this model then you could consider the reasoning process to be essentially a prior and an updating procedure.

The notion of an objective prior distribution has been strongly criticised and many have even argued that the concept is meaningless. Fine, for example, claims in [9] that

> 'The interpretation of a classical prior distribution is particularly problematic: it seems to be an objectification of subjective prior knowledge. Extracting unique quantitative probabilities from ignorance or very little prior knowledge can only harmfully obscure our ignorance. '

It seems, though, that this ignores the necessity for human beings to make objective judgements based on inadequate information by suggesting that human ignorance has no objective structure.

The idea of second order distributions in Bayesian inference has been widely discussed by, among others, de Finetti (see [10]) and Good (see [14] [13] and [11]) who comments

> 'Most of the justifications of the axioms of subjective probability assume sharp probabilities or clear-cut decisions, but there is always some vagueness and one way of trying to cope with it is to allow for the confidence that you feel in your judgements and to represent this confidence by probabilities of a higher type.'

Good goes on to suggest third and higher order probabilities are defined until

> 'the guessed expected utility of going further becomes negative if the cost is taken into account'

However, we would argue, that for modelling inexact reasoning this approach is not natural since human beings do not encounter higher order statements in a context where it is natural to make belief judgements concerning them or to make inferences directly from them. For instance, scientists do not tend to make inferences on statements expressing their degree of certainty in evidence but rather from the evidence itself. Hence, if there is any objective structure to human ignorance relating to reasoning on beliefs it should be expressible at the second order level.

Definition 21 *Let $\mathcal{V}^{(n)}$ be the algebra of Borel subsets of $V^{(n)}$*

Then we define a second order distribution to be a probability measure, $Prob^{(n)}$*, on* $\mathcal{V}^{(n)}$

Here we perceive $E(\alpha_i^{(n)})$ for $i = 1, \ldots 2^n$ to be a sequence of random variables denoted by X_i for $i = 1, \ldots 2^n$ with range $[0, 1]$ which are *a priori* dependent according to the condition $\sum_{i=1}^{2^n} X_i = 1$

A corresponding second order conditional distribution is defined as is standard by:

$$\forall S, R \in \mathcal{V}^{(n)} \textit{ such that } Prob^{(n)}(R) > 0 \ Prob^{(n)}(S|R) = \frac{Prob^{(n)}(S \cap R)}{Prob^{(n)}(R)}$$

For notational elegance if C, K are open interval probability states such that $Prob^{(n)}(V^{(n)}(K)) > 0$ then we write $Prob^{(n)}(C)$ for $Prob^{(n)}(V^{(n)}(C))$ and $Prob^{(n)}(C|K)$ for $Prob^{(n)}(V^{(n)}(C)|V^{(n)}(K))$. Also, in the sequel all integrals are lebesque integrals and the term integrable is taken to mean lebesque integrable.

We now make the following smoothness assumptions regarding $Prob^{(n)}$:

$Prob^{(n)}$ has a density function $f^{(n)}$ which is continuous on the interior of $V^{(n)}$

Hence, if C is a probability state such that $dim(V^{(n)}(C)) = 2^n - 1$ then

$$Prob^{(n)}(C) = \int_{V^{(n)}(C)} f^{(n)}(\vec{x}) \ dV^{(n)}$$

Note that, by proposition 15 , this is true of open interval probability states.

Notice that for any convex $S \in \mathcal{V}^{(n)}$ such that $dim(S) < 2^n - 1$ $Prob^{(n)}(S) = 0$. In particular, this is true of equality probability states. A possible justification is that such probability states could only be known to hold given some understanding of the meaning of the propositional variables they contain and hence they should *a priori* have probability zero

Definition 22 *Let $\mathcal{C}$ be the class of all density functions on $V^{(n)}$ for some $n > 0$ which are continuous on $Int(V^{(n)})$*

Definition 23 *Let $\mathcal{C}^j$ for $j \in \mathbb{N}$ be the class of all density functions in $\mathcal{C}$ for which all j'th order partial derivatives exist and are continuous*

In the sequel the word density will refer to density functions in $\mathcal{C}$

In this model the updating procedure will involve calculating the conditional density function relative to a knowledge base as follows.

If $f^{(n)}$ is a prior density in $\mathcal{C}$ and K is a knowledge base such that $Prob^{(n)}(K) > 0$ and $dim(V^{(n)}(K)) = 2^n - 1$ then $\forall \vec{x} \in V^{(n)}(K)$

$$f^{(n)}(\vec{x}|K) = \frac{f^{(n)}(\vec{x})}{\int_{V^{(n)}(K)} f^{(n)}(\vec{x}) dV^{(n)}}$$

so that for probability states C and K where $dim(V^{(n)}(K)) = 2^n - 1$

$$Prob^{(n)}(C|K) = \int_{V^{(n)}(K) \cap V^{(n)}(C)} f^{(n)}(\vec{x}|K) dV^{(n)}$$

As mentioned above $Prob^{(n)}(K) = 0$ for $K \in KB(L^{(n)})$ such that $dimV^{(n)}(K) < 2^n - 1$. It is often the case, however, that we wish to conditionalise on such knowledge bases . In view of this a possible extension of the definition of conditional density is as follows:

For K a knowledge base of $L^{(n)}$ such that $dim(V^{(n)}(K)) < 2^n - 1$

$$\forall \vec{x} \in V^{(n)}(K)\ f^{(n)}(\vec{x}|K) = \frac{f^{(n)}(\vec{x})}{\int_{V^{(n)}(K)} f^{(n)}(\vec{x}) dV^{(n)}(K)}$$

if $f^{(n)}$ is integrable in relative dimension on $V^{(n)}(K)$ and $\int_{V^{(n)}(K)} f^{(n)}(\vec{x}) dV^{(n)}(K) > 0$ and is left undefined otherwise.

Then for $C \in KB(L^{(n)})$

$$Prob^{(n)}(C|K) = \int_{V^{(n)}(K) \cap V^{(n)}(C)} f^{(n)}(\vec{x}|K) dV^{(n)}(K)$$

We shall refer to this extended definition as the **conditionalisation assumption**

Definition 24 *A hierarchy of second order prior distributions is a sequence* $\{Prob^{(n)}\}$*and similarly a hierarchy of prior densities is a sequence* $\{f^{(n)}\}$ *for* $n \in \mathbb{N},\ n > 0$

To summarize then we will be considering reasoning processes consisting of prior densities in $\mathcal{C}$ and an updating procedure which involves calculating the conditional density as defined above.

The notion of inference process can be linked to this model as follows. Given an open interval knowledge base K such that $Prob^{(n)}(K) > 0$ the expected value of $E(\theta)$ generated according to $f^{(n)}$ is

$$\int_{V^{(n)}(K)} \sum_{i \in S_\theta} x_i\ f^{(n)}(\vec{x}|K) dV^{(n)} = \frac{\int_{V^{(n)}(K)} \sum_{i \in S_\theta} x_i f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}(K)} f^{(n)}(\vec{x}) dV^{(n)}} = CM_K^{f^{(n)}}(\theta)$$

Similarly, given the conditionalisation assumption we have for equality knowledge bases K such that $f^{(n)}$ is integrable in relative dimension on $V^{(n)}(K)$ and $\int_{V^{(n)}(K)} f^{(n)}(\vec{x})dV^{(n)}(K) > 0$

$$\int_{V^{(n)}(K)} \sum_{i \in S_\theta} x_i \, f^{(n)}(\vec{x}|K)dV^{(n)} = \frac{\int_{V^{(n)}(K)} \sum_{i \in S_\theta} x_i f^{(n)}(\vec{x})dV^{(n)}(K)}{\int_{V^{(n)}(K)} f^{(n)}(\vec{x})dV^{(n)}(K)} = CM_K^{f^{(n)}}(\theta)$$

If $CM_K^{f^{(n)}}(\alpha_i^{(n)}) = y_i$ then $\vec{y}$ is the center of mass of the region $V^{(n)}(K)$ calculated according to $f^{(n)}$. Since $V^{(n)}(K)$ is convex $\vec{y} \in V^{(n)}(K)$ and this implies that $CM_K^{f^{(n)}}$ is a probability function on $SL^{(n)}$ consistent with K. In the sequel we will see that for some restricted classes of priors, $f^{(n)}$, $CM^{f^{(n)}}$ on equality knowledge bases is an equality inference process and that for all priors satisfying a non nullity condition $CM^{f^{(n)}}$ on open interval knowledge bases is an open interval inference process. Such inference processes are naturally referred to as Centre of Mass relative to density $f^{(n)}$.

Chapter 2

Principles on Priors

2.1 An Axiomatic Framework

The problem of choosing second order densities has parallels with the old statistical problem of defining priors for use in Bayesian inference. It has often been pointed out that for some types of Bayesian reasoning the choice of prior is relatively unimportant assuming large sample sizes. Furthermore Bayesian statisticians have, in spite of the arguments put forward by Johnson [15] and Carnap [5], tended to adopt a rather ad hoc approach to selecting priors. In the present context of inexact reasoning, however, the choice of prior is of fundamental importance and hence we feel it is necessary to develop an approach based on epistemological considerations.

The traditional response to the problem, attributed to Laplace in the case $n = 1$ (see [23] for discussion), is that in the absence of any other information the principle of insufficient reason forces us to choose $f^{(n)}$ to be the uniform distribution (i.e $Prob^{(n)}(R)$ proportional to the volume of R for $R \in \mathcal{V}^{(n)}$.) However, we shall demonstrate, at least for the more general n dimensional problem, that this is unsatisfactory.

The epistemological conditions that we propose can be separated into two categories ; **global conditions** which describe relationships between different levels of the hierarchy, and **local conditions** which are defined for fixed n. The first axioms, Renaming and Weak Renaming are local conditions.

A1:Renaming

Let $\sigma : \overline{SL^{(n)}} \longmapsto \overline{SL^{(n)}}$ be an automorphism on the Lindenbaum algebra.

For $\alpha_i^{(n)}$ $i = 1, \ldots, 2^n$ let $\sigma(\alpha_i^{(n)}) = \alpha_j^{(n)}$ where $\sigma(\overline{\alpha_i^{(n)}}) = \overline{\alpha_j^{(n)}}$.

For each such σ there is a corresponding permutation for each $\vec{x} \in V^{(n)}$ such that

$$\sigma(x_i) = x_j \iff \sigma(\alpha_i^{(n)}) = \alpha_j^{(n)}$$

Then $Prob^{(n)}$ satisfies Renaming iff for all such σ

$$Prob^{(n)}(\bigwedge_{i=1}^{2^n} E(\alpha_i^{(n)}) \in I_i) = Prob^{(n)}(\bigwedge_{i=1}^{2^n} E(\sigma(\alpha_i^{(n)})) \in I_i)$$

where I_i is a strict interval of $[0, 1]$ for $i = 1, \ldots, 2^n$. It can be easily seen that for densities, $f^{(n)}$ satisfies renaming iff

$$\forall \vec{x} \in Int(V^{(n)}) \quad f^{(n)}(\sigma(\vec{x})) = f^{(n)}(\vec{x})$$

where for all σ

$$\sigma(\vec{x}) =< \sigma(x_1), \ldots, \sigma(x_{2^n}) >$$

The justification for this principle is that the atoms merely represent possible valuations of $L^{(n)}$ each of which has *a priori* the same status. Thus the status of the prior knowledge should be invariant with respect to the ordering of the atoms.

However, consider $\theta, \varphi \in SL^{(n)}$ such that $\theta \equiv \bigvee_{i \in S_\theta} \alpha_i^{(n)}$ and $\varphi \equiv \bigvee_{i \in S_\varphi} \alpha_i^{(n)}$ where $|S_\theta| = |S_\varphi|$. Now let θ' and φ' be the simplest sentences of $L^{(n)}$, in that they contain the least number of distinct propositional variables, equivalent to θ and φ respectively. Clearly φ' and θ' may contain different numbers of propositional variables but according to the Renaming Principle the statements $\theta' \in I$ and $\varphi' \in I$, for I a strict interval of [0,1], are given equal probability. For example, given $L^{(2)} = \{p, q\}$ the statements $E(p) \in (\frac{1}{2} - \delta, \frac{1}{2} + \delta)$, $E((p \vee \neg q) \wedge (\neg p \vee q)) \in (\frac{1}{2} - \delta, \frac{1}{2} + \delta)$ have equal probability according to Renaming.

This would seem to overlook the special significance attached to the random propositional variables of the language, these being interpreted as random variables relating to the most elementary events under consideration. Therefore, it would seem natural that the complexity of a sentence of $L^{(n)}$, taken as a measure of the number of distinct propositional variables it contains, expressed in its simplest form should be taken into account when evaluating second order probabilities. In view of this an alternative axiom is as follows:

A2:Weak Renaming

$Prob^{(n)}$ satisfies **Weak Renaming** if $Prob^{(n)}$ satisfies **Renaming** where the automorphisms are restricted to those such that for all propositional variables p_i

$\sigma(\overline{p_i}) = \overline{p_j^\epsilon}$ where $\epsilon \in \{0,1\}$

A3:Marginality

For all $n > 1$ if $\alpha_i^{(n-1)}$ for $i = 1, \ldots, 2^{n-1}$ denote the atoms of $L^{(n-1)}$ then

$$Prob^{(n)}(\bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)}) \in I_i) = Prob^{(n-1)}(\bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)}) \in I_i)$$

where I_i is a strict interval of $[0,1]$ for $i = 1, \ldots, 2^{n-1}$

Expressed in terms of densities: let $\{f^{(n)}\}$ be a hierarchy of prior densities and let

$$Y_i = X_{2i-1} + X_{2i} \; for \; i = 1, \ldots, 2^{n-1}$$

If $h^{(n)}(\vec{y})$ is the marginal density function of $\vec{Y}$ induced by $f^{(n)}$ then $\{f^{(n)}\}$ is said to satisfy **Marginality** iff

$$\forall \; n > 1 \; \forall \vec{y} \in Int(V^{(n-1)}) \; f^{(n-1)}(\vec{y}) = h^{(n)}(\vec{y})$$

Note that in our enumeration the atoms of $L^{(n-1)}$ are related to those of $L^{(n)}$ by $\alpha_{2j}^{(n)} \wedge \alpha_{2j-1}^{(n)} \equiv \alpha_j^{(n-1)}$ where $\alpha_{2j-1}^{(n)} \equiv \alpha_j^{(n-1)} \wedge p_n$ and $\alpha_{2j}^{(n)} \equiv \alpha_j^{(n-1)} \wedge \neg p_n$

This axiom is motivated by the feeling that adding another propositional variable p_n to the language $L^{(n-1)}$ should not effect our knowledge of sentences that do not contain p_n. The axiom of Marginality is sufficient to guarantee this in the presence of Weak Renaming.

We can now see that the assumption of the uniform measure is unsatisfactory in that it is inconsistent with Marginality. In other words if $Prob^{(n)}$ is defined to be the uniform measure at each level n then the hierarchy of priors does not satisfy Marginality. In addition, we observe that the uniform density is a special case of the symmetric Dirichlet system of densities given by

$$\forall \vec{x} \in Int(V^{(n)}) \; d(\lambda, n)(\vec{x}) = \Gamma(\lambda)[\Gamma(\frac{\lambda}{2^n})]^{-2^n} \prod_{i=1}^{2^n} x_i^{\frac{\lambda}{2^n}-1}$$

where λ is a parameter on $(0, \infty)$ corresponding to the uniform distribution when $\lambda = 2^n$. This system of priors is frequently used in Bayesian reasoning and has been justified in this context by Johnson [15] and Carnap [5]. The inconsistency of the uniform measure with Marginality is clarified by the next result.

Proposition 25 *If* $\{d(\lambda, n)\}$ *is a hierarchy of symmetric Dirichlet priors then* $\{d(\lambda, n)\}$ *satisfies Marginality iff* $\lambda > 0$ *is a constant independent of* n.

Proof

From Wilks [26] p181 if $f^{(n)}(\vec{x}) = d(\lambda, n)(\vec{x})$ then $h^{(n)}(\vec{y}) = d(\lambda, n-1)(\vec{y})$

where $h^{(n)}$ is the marginal density defined above $\square$

A4:Non Nullity

$$Prob^{(n)}(\bigwedge_{i=1}^{2^n} E(\alpha_i^{(n)}) \in I_i) > 0$$

if

$$V^{(n)}(\{E(\alpha_i^{(n)}) \in I_i, i = 1, \ldots 2^n\}) \neq \emptyset$$

where I_i are open intervals of $(0, 1)$

In terms of densities this means that for all non empty subsets R of $V^{(n)}$ open in $D^{(n)}$ $\exists \vec{x} \in R$ such that $f^{(n)}(\vec{x}) > 0$

The justification for this principle is that any consistent open interval probability state should *a priori* have non zero probability.

A5:Strong Non Nullity

A prior density $f^{(n)}$ satisfies Strong Non Nullity if $\forall \vec{x} \in Int(V^{(n)})\ f^{(n)}(\vec{x}) > 0$

This principle is difficult to justify on epistemological grounds and should really be viewed as a smoothness assumption. It is stated at this point because of it's clear relationship to Non Nullity and for reasons of presentation.

A consequence of Strong Non Nullity is that for any $K \in KB(L^{(n)})$ such that $dim(V^{(n)}(K)) < 2^n - 1$ if $f^{(n)}$ is integrable on $V^{(n)}(K)$ then $\int_{V^{(n)}(K)} f^{(n)}(\vec{x}) dV^{(n)}(K) > 0$. In other words, provided $f^{(n)}$ is integrable on $V^{(n)}(K)$ the conditional density $f^{(n)}(\vec{x}|K)$ is defined by the conditionalisation assumption.

A problem that seems to occur regularly in relation to knowledge bases for expert systems is as follows. Suppose we have knowledge regarding specific atoms of $SL^{(n-1)}$; this is to say our knowledge base consists of constraints of the form $E(\alpha_i^{(n-1)}) \in I$ where I is an interval of $[0,1]$. If a new propositional variable p_n is then added to the language regarding which we have no information what weight should we allocate to the new atoms $p_n \wedge \alpha_i^{(n-1)}$ and $\neg p_n \wedge \alpha_i^{(n-1)}$ and what is the relationship between these and the other atoms of $L^{(n)}$? The next two axioms relate to this problem.

A6:Weak Atomic Independence

For $n \geq 2$, given the conditionalisation assumption, $f^{(n)}$ satisfies Weak Atomic

Independence iff

$$Prob^{(n)}(E(p_n|\alpha_i^{(n-1)}) \in L \;|E(\alpha_i^{(n-1)}) = t)$$

$$Prob^{(n)}(E(p_n|\alpha_i^{(n-1)}) \in L \mid \bigwedge_{j\neq i}(E(\alpha_{2j}^{(n)}) = x_{2j} \wedge E(\alpha_{2j-1}^{(n)}) = x_{2j-1}\;))$$

where L is a strict interval of $[0,1]$, $t, x_{2j}, x_{2j-1} \in (0,1)$ for $j = 1,\ldots,2^n$ subject to $\sum_{j\neq i} x_{2j} + x_{2j-1} = 1 - t$ and where both conditional probabilities are assumed to exist for all such L, t, x_{2j}, x_{2j-1} where $j = 1,\ldots,2^n$

Less formally, Weak Atomic Independence states that for all $i \in \{1,\ldots,2^{n-1}\}$ $E(\alpha_{2i}^{(n)})$ and $E(\alpha_{2i-1}^{(n)})$ are independent of the probability of the other atoms of $L^{(n)}$ up to consistency with the constraint $E(\alpha_i^{(n-1)}) = t$.

A7:Relative Ignorance The distribution $Prob^{(n)}$ satisfies Relative Ignorance if for I, J strict intervals of $[0,1]$ and $n \geq 2$

$$Prob^{(n)}(E(p_n|\alpha_i^{(n-1)}) \in J|E(\alpha_i^{(n-1)}) \in I)$$

is independent of I.

In other words the probability of p_n occurring given that $\alpha_i^{(n-1)}$ has occurred is independent of $E(\alpha_i^{(n-1)})$. The idea here is that since the constraint $E(\alpha_i^{(n-1)}) \in I$ provides no information about p_n the relative distribution of weight between the atoms $p_n \wedge \alpha_i^{(n-1)}$ and $\neg p_n \wedge \alpha_i^{(n-1)}$ should be independent of I.

2.2 Multiplicative priors

Definition 26 *$f^{(n)}$ is a multiplicative prior on $V^{(n)}$ if $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i)$ where $g_i : [0,1] \to \mathbb{R}^{\geq 0}$ for $i = 1, \ldots, 2^n$*

$f^{(n)}$ is a symmetric multiplicative prior on $V^{(n)}$ if $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$ where $g : [0,1] \to \mathbb{R}^{\geq 0}$

Notation : If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ then we write $f \propto g$ iff $\exists c \in \mathbb{R}^n$ such that $\forall \vec{x} \in \mathbb{R}^n$ $f(\vec{x}) = cg(\vec{x})$

Proposition 27 *If $f^{(n)}$ is a multiplicative prior on $V^{(n)}$ satisfying Strong Non Nullity such that*

$$\forall \vec{x} \in Int(V^{(n)}) \ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i) \ = \prod_{i=1}^{2^n} h_i(x_i)$$

then

$$\forall x \in (0,1) \ h_i(x) \ \alpha \ e^{v(x_i)} g_i(x)$$

where $\forall x \in (0,1)$ $v(x)$ satisfies Cauchy's equation and $\forall x \in (0,1)$ $v(x) + v(1-x) = k$ for some $k \in \mathbb{R}^{>0}$

Lemma 28 *If $s : (0,1) \to \mathbb{R}$, $p : (0,1) \to \mathbb{R}$ and $q : (0,1) \to \mathbb{R}$ and $\forall x, y \in (0,1)$ such that $x + y \in (0,1)$*

$$s(x+y) = p(x) + q(y)$$

then $\forall x \in (0,1)$

$$p(x) = v(x) + c_1, q(x) = v(x) + c_2, s(x) = v(x) + c_3$$

where $c_1, c_2, c_3 \in \mathbb{R}$ *and* $v : (0,1) \to \mathbb{R}$ *is such that* $\forall x, y \in (0,1)$ *for which* $x + y \in (0,1)$

$$v(x+y) = v(x) + v(y)$$

Proof

$\forall x \in (0,1)\ \forall y \in (0, 1-x)\ s(x+y) = s(y+x)$

$$\therefore\ p(x) + q(y) = q(x) + p(y)$$

$$\Rightarrow\ \forall y \in (0,1)\ q(y) = c_1 + p(y)$$

for $c_1 \in \mathbb{R} \therefore \forall x \in (0,1)\ \forall y \in (0, 1-x)$

$$s(x+y) = p(x) + p(y) + c_1$$

Now $\forall x \in (0,1)\ \forall y \in (0, 1-x)\ \forall t \in (0, 1-x-y)$

$s((x+t)+y) = s((x+y)+t) \therefore$

$$p(x+t) + p(y) = p(x+y) + p(t)$$

$\therefore\ \forall x \in (0,1)\ \forall y \in (0, 1-x)$

$$p(x+y) - p(y) = f(x)$$

where $f : (0,1) \to \mathbb{R}$

Now by the same argument as above we have $\forall x \in (0,1)\ \forall y \in (0, 1-x)$

$$f(y) + p(x) = f(x) + p(y)$$

$$\Rightarrow\ \forall x \in (0,1)\ f(x) = p(x) - c_2$$

for some $c_2 \in \mathbb{R}$

$$\therefore\ \forall x \in (0,1)\ \forall y \in (0,1-x)\ p(x+y) = p(x) + p(y) - c_2$$

$$\Rightarrow\ s(x+y) = p(x+y) + c_1 + c_2$$

That is

$$\forall x \in (0,1)\ s(x) = p(x) + c_1 + c_2$$

Now making the substitution $v(x) = p(x) - c_2$ gives

$$\forall x \in (0,1)\ \forall y \in (0,1-x)\ v(x+y) = v(x) + v(y)$$

□

Proof of Proposition 27

For all $x \in (0,1)$ let $r_i(x) = \frac{h_i(x)}{g_i(x)}$ then r_i is well defined on $(0,1)$ since by Strong Non Nullity $f^{(n)}$ is strictly positive on the interior of $V^{(n)}$

Clearly

$$\forall \vec{x} \in Int(V^{(n)})\ \prod_{i=1}^{2^n} r_i(x_i) = 1$$

Letting $y_i = x_{2i-1} + x_{2i}$ and making the change of variables to x_{2i-1}, y_i for $i = 1, \ldots, 2^{n-1}$ we obtain

$$\prod_{i=1}^{2^{n-1}} r_{2i-1}(x_{2i-1}) r_{2i}(y_i - x_{2i-1}) = 1$$

Now the right hand side is independent of x_{2i-1} for $i = 1, \ldots, 2^{n-1}$ and since for each i x_{2i-1} only appears in the product

$$r_{2i-1}(x_{2i-1}) r_{2i}(y_i - x_{2i-1})$$

it follows that $\forall x_{2i-1} \in (0, y_i) \forall y_i \in (0,1)$

$$r_{2i-1}(x_{2i-1})r_{2i}(y_i - x_{2i-1}) = r_i^*(y_i)$$

for some $r_i^* : (0,1) \to \mathbb{R}^{>0}$

Taking log's and letting $p_i = \log r_{2i},\ q_i = \log r_{2i-1},\ s_i - \log r_i^*$ we obtain that $\forall x \in (0,1)\ \forall y \in (0, 1-x)$

$$s_i(x+y) = p_i(x) + q_i(y)$$

and by lemma 28 $\forall x \in (0,1)$

$$s_i(x) = v_i(x) + c_1,\ p_i(x) = v_i(x) + c_2,\ q_i(x) = v_i(x) + c_3$$

where $c_1, c_2, c_3 \in \mathbb{R}$ and $v_i : (0,1) \to \mathbb{R}$ such that $\forall x \in (0,1)\ \forall y \in (0, 1-x)$

$$v_i(x+y) = v_i(x) + v_i(y)$$

Now for $j \in \{2, \ldots, 2^{n-1}\}$ let $x_1 + x_{2j-1} = y_1$, $x_2 + x_{2j} = y_j$ and $x_{2i} + x_{2i-1} = y_i$ for $i = 2, \ldots, 2^{n-1}\ i \neq j$ and make the change of variables to x_1, x_2, y_1, y_j and x_{2i-1}, y_i for $i = 2, \ldots, 2^{n-1}\ i \neq j$. Then considering

$$r_1(x_1)r_{2j-1}(y_1 - x_1)$$

as above for $j = 2, \ldots, 2^{n-1}$ we obtain that $\forall j \in \{1, \ldots, 2^n\}\ \forall x \in (0,1)$

$$v_j(x) = v_1(x) + \lambda_j$$

for some $\lambda_j \in \mathbb{R}$

Further since $\prod_{i=1}^{2^{n-1}} r_i(x_i) = 1$ it follows that

$$\forall x \in (0,1)\ v_1(x) + v_1(1-x) = k$$

for $k \in \mathbb{R}^{>0}$ $\square$

Corollary 29 *If* $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i) = \prod_{i=1}^{2^n} h_i(x_i)$ *then* $h_i(x)\ \alpha\ e^{cx} g_i(x)$ $\forall x \in \mathbb{Q} \cap (0,1)$

Proof

Let $v(x) + v(y) = v(x+y)$ for $x, y, x+y \in (0,1)$ then by induction on n for $n \in \mathbb{N}$ $n > 0$ we have that

$$\forall x < \frac{1}{n}\ v(nx) = nv(x)$$

Letting $x = \frac{1}{2n}$ gives that $v(\frac{1}{2}) = nv(\frac{1}{2n})$ which implies that $v(\frac{1}{2n}) = c\frac{1}{2n}$ for some constant $c \in \mathbb{R}$

Hence, for $m < n$ $m, n \in \mathbb{N}$

$$v(2m\frac{1}{2n}) = 2mv(\frac{1}{2n})$$

therefore,

$$v(\frac{m}{n}) = 2mc\frac{1}{2n} = c\frac{m}{n}$$

$\square$

The above result illustrates that multiplicative representations of priors are not unique. That is, for any prior $f^{(n)}$ where $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i)$ it is also the case that $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} e^{v(x_i)} g_i(x_i)$ where $v : (0,1) \rightarrow \mathbb{R}$ and $\forall x \in (0,1)$ $\forall y \in (0, 1-x)$ satisfies $v(x+y) = v(x) + v(y)$ and $v(x) + v(1-x) = 1$

Corollary 30 *Assuming the Axiom of Choice, if $f^{(n)}$ is a multiplicative prior continuous on the interior of $V^{(n)}$ satisfying Strong Non Nullity then there exists functions*

$$g_i : [0,1] \longrightarrow \mathbb{R}^{\geq 0} \ for \ i = 1, \ldots, 2^n$$

where g_i is not continuous on $(0,1)$ such that

$$\forall \vec{x} \in Int(V^{(n)}) \ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i)$$

Proof

Suppose that $\forall \vec{x} \in Int(V^{(n)}) \ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i)$ where g_i is continuous on $(0,1)$.

Now by Zorn's Lemma every vector space has a basis. Therefore, let B be a Hammel basis for the vector space $\mathbb{R}$ over $\mathbb{Q}$. Let $b_0 \in B$. Then for $r \in \mathbb{R}$ there are unique q_0, q_i in $\mathbb{Q}$ such that

$$r = q_0 b_0 + \sum_{b_i \in B - \{b_0\}} q_i b_i$$

Now set $v(r) = q_0$ then

$$\forall s, r \in \mathbb{R} \ v(r+s) = v(r) + v(s)$$

but it is not the case that $\forall r \in \mathbb{R} \ v(r) = mr$ for some constant $m \in \mathbb{R}$ since $v(b_0) = 1$ and $v(b) = 0$ for $b \in B - \{b_0\}$.

then

$$\forall \vec{x} \in Int(V^{(n)}) \ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} \frac{e^{v(x_i)} g_i(x_i)}{e^{v(1)}}$$

and by corollary 29 $e^{v(x)}g_i(x)$ is not continuous on $(0,1)$ $\square$

In view of the above corollary, in the sequel we shall consider only the following categories of multiplicative priors.

Definition 31 *(i) Let $\mathcal{C}_M$ be the class of multiplicative priors $\prod_{i=1}^{2^n} g_i(x_i)$ in $\mathcal{C}$ where each g_i is continuous on $(0,1)$.*

(ii) Let $\mathcal{C}_M^j$ be the class of multiplicative priors $\prod_{i=1}^{2^n} g_i(x_i)$ in $\mathcal{C}$ where the j'th derivatives of each g_i exist and are continuous on $(0,1)$.

(iii) Let $\mathcal{C}_M^\infty$ be the class of multiplicative priors $\prod_{i=1}^{2^n} g_i(x_i)$ in $\mathcal{C}$ where all the derivatives of each g_i exist and are continuous on $(0,1)$ and $\exists a_i \in (0,1)$ such that $\forall x \in (0,1)$ the Taylors series of $g_i(x)$ about a_i converges.

Definition 32 *If $g_i : [0,1] \to \mathbb{R}^{\geq 0}$ and $g_j : [0,1] \to \mathbb{R}^{\geq 0}$ then $\forall y \in (0,1)$*

$$g_i * g_j(y) = \int_0^y g_i(x)g_j(y-x)dx$$

provided the integral exists and is undefined otherwise

By a simple argument involving a change of variables it can be seen that $\forall y \in (0,1)$ $g_i * g_j(y) = g_j * g_i(y)$. In fact $*$ is the convolution product for functions on $\mathbb{R}^{\geq 0}$. For a discussion on convolution products of distributions see Feller [8] or for a more in depth discussion of the convolution product see Kecs [16] .

Definition 33 *For $n \in \mathbb{N}$ $y \in (0,1)$*

(i)

$$SP^{(n)} \equiv \ \{\vec{x} \in [0,1]^n | \sum_{i=1}^{n} x_i = 1\}$$

(ii)

$$Int(SP^{(n)}) \equiv \{\vec{x} \in SP^{(n)} | x_i > 0 \ i = 1, \ldots, n\}$$

(iii)

$$SEG^{(n)}(y) \equiv \ \{\vec{t} \in [0,1]^n | \sum_{i=1}^{n} t_i \leq y\}$$

(iv)

$$ISEG^{(n)}(y) \equiv \ \{\vec{t} \in (0,1)^n | \sum_{i=1}^{n} t_i < y\}$$

Definition 34 *For* $m \in \mathbb{N}$ $m \geq 1$

$$\forall y \in [0,1] \ g^{*m}(y) = \int_{SEG^{(m-1)}(y)} (\prod_{i=1}^{m-1} g(x_i)) g(y - \sum_{i=1}^{m-1} x_i) d(\vec{x})$$

if this integral exists and is left undefined otherwise.

Proposition 35 *For all symmetric multiplicative priors* $f^{(n)} \in \mathcal{C}_M$ *satisfying Non Nullity where* $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$ *we have that for* $m \geq 2$ $m \in \mathbb{N}$ $g^{*m}(y)$ *exists* $\forall y \in (0,1)$, g^{*m} *is continuous on* $(0,1)$ *and is integrable on* $[0, \epsilon]$ $\forall \epsilon \in (0,1)$

Lemma 36 *If* $g : [0,1] \to \mathbb{R}^{\geq 0}$ *is continuous on* $(0,1)$ *and* $\exists \ y \in (0,1)$ *and* $n \in \mathbb{N}$ *,* $n > 0$ *such that* $\forall \vec{t} \in \ ISEG^{(n)}(y)$ $(\prod_{i=1}^{n} g(t_i)) g(y - \sum_{i=1}^{n} t_i) = 0$ *then there is some open interval* I *of* $(0,1)$ *such that* $\forall t \in I$ $g(t) = 0$

Proof

Suppose there exists

$x_1 \in (0, y)$, $x_2 \in (0, y - x_1)$, ... , $x_{n-1} \in (0, y - \sum_{i=1}^{n-2} x_i)$

such that $g(x_i) > 0$ for $i = 1, \ldots, n-1$ otherwise the result follows trivially

We have then that

$$\forall t \in (0, y - \sum_{i=1}^{n-1} x_i) \;\; g(t)g(y - \sum_{i=1}^{n-1} x_i - t) = 0$$

Now suppose $\exists x_n \in (0, y - \sum_{i=1}^{n-1} x_i)$ such that $g(x_n) > 0$ then by the continuity of g on $(0,1)$ there exists $0 \leq a < x_n < b \leq \;\; y - \sum_{i=1}^{n-1} x_i$ such that $\forall t \in (a,b) \; g(t) > 0$

$$\therefore \;\; \forall t \in (a,b) \; g(y - \sum_{i=1}^{n-1} x_i - t) = 0$$

In other words

$$\forall t \in (y - \sum_{i=1}^{n-1} x_i - b, y - \sum_{i=1}^{n-1} x_i - a) \;\; g(t) = 0$$

$\square$

Lemma 37 *If $f^{(n)} \in \mathcal{C}_M$ is a symmetric multiplicative prior satisfying Non Nullity such that $\forall \vec{x} \in V^{(n)} \;\; f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$ then $g * g(y)$ exists for almost all $y \in (0,1)$*

Proof

Since $f^{(n)}$ is a density function on $V^{(n)}$ we have that $\int_{V^{(n)}} f^{(n)} dV^{(n)}$ exists

Now let $g' : \mathbb{R} \to \mathbb{R}$ be defined as follows:

$$g'(x) = g(x) \ for \ x \in [0,1]$$

$$g'(x) = 0 \ otherwise$$

then

$$\int_{V^{(n)}} f^{(n)} dV^{(n)} = \int_{\mathbb{R}\times\mathbb{R}^{2^n-2}} g'(x_1)g'(y-x_1)(\prod_{i=2}^{2^n-2} g'(x_i))g'(1-y-\sum_{i=2}^{2^n-2} x_i)d(\vec{x},y)$$

Now by Fubini's theorem (see [2]) $\exists T \subseteq [0,1]$ such that T has 1 dimensional Lebesque measure 0 and $\forall y \in [0,1] - T$

$$\int_{\mathbb{R}^{2^n-2}} g'(x_1)g'(y-x_1)(\prod_{i=2}^{2^n-2} g'(x_i))g'(1-y-\sum_{i=2}^{2^n-2} x_i)d(\vec{x})$$

exists

Let $y \in [0,1] - T$ then

$$\int_{\mathbb{R}^{2^n-3}\times\mathbb{R}} (\prod_{i=1}^{2^n-3} g'(t_i))g'(1-y-\sum_{i=1}^{2^n-3} t_i)g'(x)g'(y-x)d(x,\vec{t})$$

exists

Again by Fubini's theorem $\exists S \subseteq \mathbb{R}^{2^n-3}$ such that S has $2^n - 3$ dimensional Lebesque measure 0 and $\forall \vec{t} \in \mathbb{R}^{2^n-3} - S$

$$\int_{\mathbb{R}} (\prod_{i=1}^{2^n-3} g'(t_i))g'(1-y-\sum_{i=1}^{2^n-3} t_i)g'(x)g'(y-x)dx$$

exists

$\therefore$

$$\int_0^y (\prod_{i=1}^{2^n-3} g(t_i))g(1-y-\sum_{i=1}^{2^n-3} t_i)g(x)g(y-x)dx$$

exists for all $\vec{t} \in ISEG^{(2^n-3)}(1-y) - S'$ where $S' \equiv S \cap ISEG^{(2^n-3)}(1-y)$

Now suppose $\forall \vec{t} \in ISEG^{(2^n-3)}(1-y) - S'$

$$(\prod_{i=1}^{2^n-3} g(t_i))g(1-y-\sum_{i=1}^{2^n-3} t_i) = 0$$

and further suppose $\exists \vec{t} \in S'$ such that

$$(\prod_{i=1}^{2^n-3} g(t_i))g(1-y-\sum_{i=1}^{2^n-3} t_i) > 0$$

then clearly $g(t_i) > 0$ for $i = 1, \ldots, 2^n - 3$ and by continuity of g on $(0,1)$ it follows that g is strictly positive on some open interval (a_i, b_i) where $0 \le a_i < t_i < b_i \le 1$ for $i = 1, \ldots, 2^n - 3$

Now let

$$I \equiv \{\vec{t} \in ISEG^{(2^n-3)}(1-y) |\ t_i \in (a_i, b_i)\ i = 1, \ldots, 2^n - 3\}$$

then

$$\forall \vec{t} \in I \quad (\prod_{i=1}^{2^n-3} g(t_i))g(1-y-\sum_{i=1}^{2^n-3} t_i) > 0$$

$\therefore$ $I \subseteq S'$ but this is a contradiction since the $2^n - 3$ dimensional Lebesgue measure of I is greater than 0

Hence, $\forall \vec{t} \in ISEG^{(2^n-3)}(1-y)$ $(\prod_{i=1}^{2^n-3} g(t_i))g(1-y-\sum_{i=1}^{2^n-3} t_i) = 0$

$\therefore$ by lemma 36 g is zero on some open interval of $(0,1)$ contradicting Non Nullity

$\therefore\ \exists \vec{t} \in ISEG^{(2^n-3)}(1-y) - S'$ such that

$$(\prod_{i=1}^{2^n-3} g(t_i))g(1-y-\sum_{i=1}^{2^n-3} t_i) > 0$$

$$\therefore\ \int_0^y g(x)g(y-x)dx$$

exists $\forall y \in [0,1] - T$ □

Lemma 38 *If $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ are integrable on $\mathbb{R}$ then $g * h$ is integrable on $\mathbb{R}$*

Proof

See de Barra [3] page 214 □

Lemma 39 *If $h : [0,1] \to \mathbb{R}^{\geq 0}$ and $g : [0,1] \to \mathbb{R}^{\geq 0}$ are continuous on $(0,1)$ and integrable on $[0,\epsilon]$ $\forall \epsilon \in (0,1)$ then $\forall x \in (0,1)$ $g * h(x)$ exists and is continuous at x*

Proof

Now $\forall y \in (\epsilon,1)$ $g(y-x)$ is continuous at x $\forall x \in [0,\epsilon]$

Therefore, since h is integrable on $[0,\epsilon]$ $h(x)g(y-x)$ is integrable on $[0,\epsilon]$

Similarly $\forall y \in (\epsilon,1)$ $h(y-x)$ is continuous at x $\forall x \in [0,y-\epsilon]$

Therefore, since g is integrable on $[0, y-\epsilon]$ we have that $g(x)h(y-x)$ is integrable on $[0, y-\epsilon]$

$\Rightarrow g(y-x)h(x)$ is integrable on $[\epsilon, y]$

Hence $\forall y \in (\epsilon, 1)$

$$\int_0^\epsilon h(x)g(y-x)dx + \int_\epsilon^y h(x)g(y-x)dx = \int_0^y h(x)g(y-x)dx$$

exists and since ϵ can be arbitrarily close to zero we have that

$$\forall y \in (0,1) \quad \int_0^y h(x)g(y-x)dx$$

exists

Now $\forall \epsilon \in (0,1)$ $\forall y \in (2\epsilon, 1)$

$$\int_0^y h(x)g(y-x)dx = \int_\epsilon^{y-\epsilon} h(x)g(y-x)dx + \int_0^\epsilon h(x)g(y-x)dx + \int_{y-\epsilon}^y h(x)g(y-x)dx$$

$$= \int_\epsilon^{y-\epsilon} h(x)g(y-x)dx + \int_0^\epsilon h(x)g(y-x)dx + \int_0^\epsilon g(x)h(y-x)dx$$

Now $\forall \epsilon \in (0,1)$ $\forall y \in (\epsilon, 1]$ $h(x)$ and $g(x)$ are uniformly continuous on $[\epsilon, y-\epsilon]$ hence it can easily be seen that $\int_\epsilon^{y-\epsilon} h(x)g(y-x)dx$ is continuous at y $\forall y \in (2\epsilon, 1)$

Further $\forall \epsilon \in (0,1)$ $\forall y \in (\epsilon, 1)$ $g(x)$ is uniformly continuous on $[y-\epsilon, y]$

Therefore, $\forall \Delta > 0$ $\exists \alpha > 0$ such that $\forall \delta$ where $|\delta| < \alpha$ we have that $\forall x \in [0, \epsilon]$

$$|g(y-x) - g(y+\delta-x)| < \Delta$$

Then since h is integrable on $[0,\epsilon]$ it follows that

$$|\int_0^\epsilon h(x)g(y-x)dx - \int_0^\epsilon h(x)g(y+\delta-x)dx| = |\int_0^\epsilon h(x)(g(y-x)-g(y+\delta-x))dx|$$

$$\leq \int_0^\epsilon h(x)|g(y-x) - g(y+\delta-x)|dx \leq \Delta \int_0^\epsilon h(x)dx$$

hence $\int_0^\epsilon h(x)g(y-x)dx$ is continuous at y $\forall y \in (2\epsilon, 1)$

Similarly we can show that $\int_0^\epsilon g(x)h(y-x)dx$ is continuous at y $\forall y \in (2\epsilon, 1)$

Hence $\forall y \in (2\epsilon, 1)$ $\int_0^y h(x)g(y-x)dx$ is continuous at y and since ϵ can be taken to be arbitrarily close to zero we have the required result $\square$

Proof of Proposition 35

We prove the result by induction on m

For $m = 2$ suppose that $\exists\epsilon \in (0,1)$ such that g is not integrable on $[0,\epsilon]$. Now since g is continuous on $(0,1)$ it follows that $g(x)$ is unbounded as x tends to zero. Hence, for all $\epsilon' \in (0,1)$ g is not integrable on $[0,\epsilon']$. Also then, by continuity of g on $(0,1)$ it follows that $\exists\delta \in (0,1)$ such that g is strictly positive on $(0,\delta)$. Therefore, for ϵ such that $\epsilon < y < \delta$ $g(y-x)$ is strictly positive for $x \in (0,\epsilon)$. Hence, $g(x)g(y-x)$ is not integrable on $[0,\epsilon]$. This implies that $\forall y \in (0,\delta)$ $\int_0^y g(x)g(y-x)dx$ does not exist which contradicts proposition 35 since $(0,\delta)$ has 1 dimensional Lebesgue measure greater than 0

Therefore, $\forall \epsilon \in (0,1)$ g is integrable on $[0,\epsilon]$

Then by lemma 39 $g * g(y)$ exists $\forall y \in (0,1)$, is continuous on $(0,1)$ and by lemma 38 $g * g$ is integrable on $[0,\epsilon]$ $\forall \epsilon \in (0,1)$.

Suppose, that $g^{*\ m-1}$ exists $\forall y \in (0,1)$, is continuous on $(0,1)$ and integrable on $[0,\epsilon]$ $\forall \epsilon \in (0,1)$. Then by lemma 39 $g^{*m}(y)$ exists $\forall y \in (0,1)$ and is continuous on $(0,1)$. Also, by lemma 38 g^{*m} is integrable on $[0,\epsilon]$ $\forall \epsilon \in (0,1)$. $\square$

Corollary 40 *Let $f^{(n)} \in \mathcal{C}_M$. Then for random variables of the form $Z_i = \sum_{j \in T_i} X_j$ where $i = 1,\ldots,m$ for $0 < m < 2^n - 1$, $T_i \cap T_j \equiv \emptyset$ for $i \neq j$, $\bigcup_{i=1}^m T_i \equiv \{1,\ldots,2^n\}$ we have that $\forall \vec{z} \in Int(SP^{(m)})$ the marginal density of $\vec{Z}$ given by*

$$\int_{V^{(n)}(K_{\vec{z}})} f^{(n)}(\vec{x}) dV^{(n)}(K_{\vec{z}})$$

where

$$K_{\vec{z}} \equiv \{\sum_{j \in T_i} E(\alpha_j^{(n)}) = z_i | i = 1,\ldots,m\}$$

exists and is continuous on $Int(SP^{(m)})$

Proof

By proposition 35

$$\int_{V^{(n)}(K_{\vec{z}})} f^{(n)}(\vec{x}) dV^{(n)}(K_{\vec{z}}) = \prod_{i=1}^{m} g^{*\ |T_i|-1}(z_i)$$

which exists for all $\vec{z} \in Int(SP^{(m)})$ and is continuous on $Int(SP^{(m)})$ $\square$

Lemma 41 *If $\{f^{(n)}\}$ is a hierarchy of symmetric multiplicative priors in $\mathcal{C}$ satisfying Marginality such that $\forall n \geq 1$ $f^{(n)}$ satisfies Strong Non Nullity and $\forall \vec{x} \in V^{(n)}$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g^{(n)}(x_i)$ then $\forall y \in (0,1)$ $g^{(n)} * g^{(n)}(y)$ exists and $\forall \vec{y} \in Int(V^{(n-1)})$ $f^{(n-1)}(\vec{y}) = \prod_{i=1}^{2^{n-1}} g^{(n)} * g^{(n)}(y_i)$*

Proof

Since $\forall n \geq 1$ $f^{(n)} \in \mathcal{C}$ we have by Marginality that

$$\forall \vec{y} \in Int(V^{(n-1)})\ f^{(n-1)}(\vec{y}) = \int_{[0,y_1]\times\ldots\times[0,y_{2^{n-1}}]} f^{(n)}(x_1, y_1 - x_1, \ldots, x_{2^n-1}, y_{2^{n-1}} - x_{2^n-1}) d(\vec{x})$$

$$= \int_{[0,y_1]\times\ldots\times[0,y_{2^{n-1}}]} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i) g^{(n)}(y_i - t_i) d(\vec{t})$$

exists.

Then by Fubini's theorem $\forall \vec{t} \in [0,y_1] \times \ldots \times [0, y_{2^{n-1}-1}] - S$ for some S with $2^{n-1} - 1$ dimensional lebesgue measure 0

$$\int_0^{y_{2^{n-1}}} \left(\prod_{i=1}^{2^{n-1}-1} g^{(n)}(t_i) g^{(n)}(y_i - t_i)\right) g^{(n)}(t_{2^{n-1}}) g^{(n)}(y_{2^{n-1}} - t_{2^{n-1}}) dt_{2^{n-1}}$$

exists which implies by Strong Non Nullity that

$$\int_0^{y_{2^{n-1}}} g^{(n)}(t) g^{(n)}(y_{2^{n-1}} - t) dt$$

exists

Therefore, clearly

$$\forall \vec{y} \in Int(V^{(n-1)})\ f^{(n-1)}(\vec{y}) = \prod_{i=1}^{2^{n-1}} g^{(n)} * g^{(n)}(y_i)$$

$\square$

From this lemma it follows that for a hierarchy of priors satisfying Marginality and Strong Non Nullity if $f^{(n)}$ is a symmetric multiplicative prior then $\forall\ 0 < m < n\ f^{(m)}$ is a symmetric multiplicative prior and further, $\forall\ 0 < m < n\ \forall \vec{y} \in V^{(m)}\ f^{(m)}(\vec{y}) = \prod_{i=1}^{2^m} g^{n-m\ *}(y_i)$. Clearly by proposition 35 the same is true for hierarchies of priors in $\mathcal{C}_M$ satisfying Marginality and Non Nullity

The next section gives a number of results relating to the axioms A1 to A7

2.3 Multiplicativity and Dirichlet Priors

The next result illustrates the connection between multiplicative priors and Weak Atomic Independence.

Theorem 42 *Let $\{f^n\}$ be a hierarchy of prior densities satisfying Marginality. For all $n \geq 1$ let $f^{(n)}$ be a density in $\mathcal{C}$ which satisfies Weak Renaming and Strong Non Nullity .Then given the conditionalisation assumption the following are equivalent*

(i) $\forall\ n\ \geq 2\ f^{(n)}$ satisfies Weak Atomic Independence

(ii) $\forall\ n\ \geq 1\ f^{(n)}$ is symmetric multiplicative prior such that $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$ where $\forall x \in (0,1)\ g^{*\ 2^n-2}(x)$ exists

Lemma 43 *If $\{f^{(n)}\}$ satisfies Marginality and Strong Non Nullity and $\forall n > 0\ f^{(n)} \in \mathcal{C}$ then given the conditionalisation assumption we have that*

$\forall n > 0$ $f^{(n)}$ *satisfies Weak Atomic Independence iff* $\forall n > 0$ $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i)$ *for some* $g_i : [0,1] \to \mathbb{R}$

Proof

By Strong Non Nullity and Weak Atomic Independence we have for all points $\vec{x}, \vec{z} \in Int(V^{(n)})$ such that $x_1 + x_2 = z_1 + z_2 = t$ and $\forall \Delta \in (0,1)$

$$Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in (\frac{x_1}{t} - \Delta, \frac{x_1}{t} + \Delta)| \bigwedge_{j=3}^{2^n} E(\alpha_j^{(n)}) = x_j)$$

$$\times Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in (\frac{z_1}{t} - \Delta, \frac{z_1}{t} + \Delta)| \bigwedge_{j=3}^{2^n} E(\alpha_j^{(n)}) = z_j)$$

$$= Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in (\frac{x_1}{t} - \Delta, \frac{x_1}{t} + \Delta)| \bigwedge_{j=3}^{2^n} E(\alpha_j^{(n)}) = z_j)$$

$$\times Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in (\frac{z_1}{t} - \Delta, \frac{z_1}{t} + \Delta)| \bigwedge_{j=3}^{2^n} E(\alpha_j^{(n)}) = x_j)$$

$\Rightarrow$

$$\frac{\int_{x_1-\Delta t}^{x_1+\Delta t} f^{(n)}(s_1, t-s_1, x_3, \ldots, x_{2^n})ds}{\int_{x_1-\Delta t}^{x_1+\Delta t} f^{(n)}(s_1, t-s_1, z_3, \ldots, z_{2^n})ds}$$

$$= \frac{\int_{z_1-\Delta t}^{z_1+\Delta t} f^{(n)}(s_1, t-s_1, x_3, \ldots, x_{2^n})ds}{\int_{z_1-\Delta t}^{z_1+\Delta t} f^{(n)}(s_1, t-s_1, z_3, \ldots, z_{2^n})ds}$$

Now since $f^{(n)} \in \mathcal{C}$ the functions of s_1, $f^{(n)}(s_1, t-s_1, x_3 \ldots, x_{2^n})$ and $f^{(n)}(s_1, t-s_1, z_3 \ldots, z_{2^n})$ are continuous on $(x_1-\Delta t, x_1+\Delta t)$ and $(z_1-\Delta t, z_1+\Delta t)$. Therefore, taking the limit as Δ tends to zero we obtain

$$f^{(n)}(x_1, t-x_1, x_3, \ldots, x_{2^n-1}, 1-t-\sum_{i=3}^{2^n-1} x_i)$$

$$\times f^{(n)}(z_1, t-z_1, z_3, \ldots, z_{2^n-1}, 1-t-\sum_{i=3}^{2^n-1} z_i)$$

$$= f^{(n)}(x_1, t-x_1, z_3, \ldots, z_{2^n-1}, 1-t-\sum_{i=3}^{2^n-1} z_i)$$

$$\times f^{(n)}(z_1, t-z_1, x_3, \ldots, x_{2^n-1}, 1-t-\sum_{i=3}^{2^n-1} x_i)$$

Fixing $z_1, z_3, \ldots, z_{2^{n-1}}$ the above equality holds for all $\vec{x} \in Int(V^{(n)})$ such that $z_1 < x_1 + x_2 < 1 - \sum_{i=3}^{2^{n-1}} z_i$.

Furthermore, we can pick $z_1, z_3, \ldots, z_{2^{n-1}}$ to be arbitrarily close to zero

$\therefore\ \forall \vec{x} \in Int(V^{(n)})$

$$f^{(n)}(x_1, t-x_1, x_3, \ldots, x_{2^n-1}, 1-t-\sum_{i=3}^{2^n-1} x_i) = U_1(x_1, t-x_1)V_1(x_3, \ldots, x_{2^n-1}, 1-t-\sum_{i=3}^{2^n-1} x_i)$$

Now for permutations σ of $\{1, \ldots, 2^{n-1}\}$ where $\sigma(1) = j$ for $j = 2, \ldots, 2^{n-1}$ repeat the above argument substituting $\alpha_{\sigma(1)}^{(n-1)}$ for $\alpha_1^{(n-1)}$, $\alpha_{2\sigma(i)}^{(n)}$ for $\alpha_{2i}^{(n)}$, $\alpha_{2\sigma(i)-1}^{(n)}$ for $\alpha_{2i-1}^{(n)}$, $x_{2\sigma(i)}$ for x_{2i} and $x_{2\sigma(i)-1}$ for x_{2i-1} where $i = 1, \ldots, 2^{n-1}$. This gives that

$$\forall \vec{x} \in Int(V^{(n)})\ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^{n-1}} U_i(x_{2i-1}, x_{2i})$$

Therefore, by Marginality we have the required result for the $n-1'th$ level in the hierarchy. $\square$

Lemma 44 *If $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i)$ is a multiplicative prior satisfying Weak Renaming and Strong Non Nullity then $f^{(n)}$ is a symmetric multiplicative prior on $V^{(n)}$.*

Proof

For any pair of atoms of $L^{(n)}$, $\alpha_i^{(n)}$ and $\alpha_j^{(n)}$ where $i \neq j$ $i, j \in \{1, \ldots, 2^n\}$, there is an automorphism σ on $\overline{SL^{(n)}}$ such that $\forall i \in \{1, \ldots, n\}$ $\sigma(\overline{p_i}) = \overline{p_j^{\epsilon}}$ where $\epsilon \in \{0, 1\}$ and $\sigma(\bar{\alpha}_i^{(n)}) = \bar{\alpha}_j^{(n)}$

Then by Weak Renaming $\forall \vec{x} \in Int(V^{(n)})$

$$g_i(x_i)g_j(x_j) \prod_{k \neq i, k \neq j} g_k(x_k) = g_i(x_j)g_j(x_i) \prod_{k \neq i, k \neq j} g_k(\sigma(x_k))$$

Therefore by proposition 27 $\forall x \in (0, 1)$

$$g_i(x) = c_1 e^{v(x)} g_j(x)$$

and

$$g_j(x) = c_2 e^{v(x)} g_i(x)$$

for $c_1, c_2 \in \mathbb{R}$ and $v : (0, 1) \to \mathbb{R}$ satisfying Cauchy's equation

$\therefore \forall x \in (0, 1)$

$$g_i(x) = c_1 c_2 e^{2v(x)} g_i(x)$$

Now by Strong Non Nullity $\forall x \in (0, 1)$ $g_i(x) > 0 \Rightarrow$

$$e^{2v(x)} = \frac{1}{c_1, c_2} \Rightarrow v(x) = \frac{1}{2} \log(\frac{1}{c_1 c_2})$$

$$\therefore \forall x \in (0,1)\ g_i(x) \propto g_j(x)$$

$\square$

Proof of Theorem 42

$((i) \Rightarrow (ii))$

It follows immediately from lemma 43 and lemma 44 that $f^{(n)}$ is a symmetric multiplicative prior. In other words there is some $g : [0,1] \to \mathbb{R}^{\geq 0}$ such that $\forall \vec{x} \in Int(V^{(n)})\ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$.

Since $\forall t \in (0,1)$ it is assumed that

$$Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in L \mid E(\alpha_1^{(n-1)}) = 1-t)$$

is defined we have that $\forall t \in (0,1)$

$$\int_{[0,1-t]\times SEG^{(2^n-3)}(t)} g(s)g(1-t-s)(\prod_{i=3}^{2^n-1} g(x_i))g(t-\sum_{i=3}^{2^n-1} x_i)d(\vec{x})ds$$

exists.

Therefore, by Strong Non Nullity and Fubini's theorem

$$\int_{SEG^{(2^n-3)}(t)} (\prod_{i=3}^{2^n-1} g(x_i))g(t-\sum_{i=3}^{2^n-1} x_i)d(\vec{x})$$

exists

$((ii) \Rightarrow (i))$

By lemma 41 $\forall y \in (0,1)\ g * g(y)$ exists

Let $\vec{x} \in V^{(n)}$ be such that $x_1 + x_2 = t$ and w.l.o.g consider

$$Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in L| \bigwedge_{j=3}^{2^n} E(\alpha_j^{(n)}) = x_j)$$

$$= \frac{\int_{tL} g(x_1)g(t-x_1)dx_1}{\int_0^t g(x_1)g(t-x_1)dx_1}$$

$$= \frac{\int_{tL} g(x_1)g(t-x_1)dx_1 g^{*\ 2^n-2}(1-t)}{\int_0^t g(x_1)g(t-x_1)dx_1 g^{*\ 2^n-2}(1-t)}$$

$$= Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in L|\ E(\alpha_1^{(n-1)}) = t)$$

□

Corollary 45 *Let $\{f^{(n)}\}$ be a hierarchy of prior densities satisfying Marginality such that for all $n > 0$ $f^{(n)}$ is a density in $\mathcal{C}$ which satisfies Strong Non Nullity, Weak Renaming and Weak Atomic Independence then given the conditionalisation assumption we have that for all $n > 0$ $f^{(n)}$ satisfies Renaming.*

Proof

Follows immediately from Theorem 42 since any symmetric multiplicative prior satisfies Renaming □

We now give a result demonstrating the equivalence, for priors in $\mathcal{C}_M^2$, between accepting the axiom of Relative Ignorance and restricting attention to symmetric Dirichlet priors.

Lemma 46 *If $\{f^{(n)}\}$ is a hierarchy of priors satisfying Marginality and $\forall n > 0$ $f^{(n)} \in \mathcal{C}_M$ and $f^{(n)}$ satisfies Non Nullity then $\forall n > 0$ $f^{(n)}$ satisfies Strong Non Nullity.*

Proof

Since $f^{(n)} \in \mathcal{C}_M$ $\forall \vec{x} \in Int(V^{(n)}) f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i(x_i)$ where $g_i : [0,1] \to \mathbb{R}^{\geq 0}$ is continuous on $(0,1)$ for $i = 1, \ldots, 2^n$

By Marginality

$$\forall \vec{y} \in Int(V^{(n-1)})\ f^{(n-1)}(\vec{y}) = \int_{[0,y_1]\times\ldots\times[o,y_{2^{n-1}}]} \prod_{i=1}^{2^{n-1}} g_{2i-1}(x_i) g_{2i}(y_i - x_i) d(\vec{x})$$

Now suppose $\exists \vec{y} \in Int(V^{(n-1)})$ such that $f^{(n-1)}(\vec{y}) = 0$ then by the continuity of g_i on $(0,1)$ for $i = 1, \ldots, 2^n$ we have that

$$\forall \vec{x} \in (0, y_1) \times \ldots \times (0, y_{2^{n-1}}) \prod_{i=1}^{2^{n-1}} g_{2i-1}(x_i) g_{2i}(y_i - x_i) = 0$$

Therefore $\exists i \in \{1, \ldots, 2^{n-1}\}$ such that

$$\forall x \in (0, y_i)\ g_{2i-1}(x) g_{2i}(y_i - x) = 0$$

Now by Non Nullity $\exists t \in (0, y_i)$ such that $g_{2i-1}(t) > 0$. Then by continuity of g_{2i-1} on $(0,1)$ $g_{2i-1}(x) > 0$ for all x in some open interval (a,b) where $a < t < b < y_i$. This implies that $\forall x \in (y_i - b, y_i - a)\ g_{2i}(x) = 0$ contradicting Non Nullity. $\square$

Theorem 47 *If $f^{(n)}$ is in $\mathcal{C}_M^2$ and satisfies Strong Non Nullity and Weak Renaming then, $f^{(n)}$ satisfies Relative Ignorance iff $f^{(n)}$ is a symmetric Dirichlet prior.*

Lemma 48 *If $g : [0,1] \to \mathbb{R}^{\geq 0}$ is a twice differentiable function on $(0,1)$ such that $\forall x \in (0,1)$ $g(x) > 0$ and*

$$g(xy).g(x(1-y)) = G(x).g(\frac{y}{2}).g(\frac{1-y}{2})$$

where $G : (0,1) \to \mathbb{R}^{\geq 0}$ *is a twice differentiable function strictly positive on* $(0,1)$ *then*

$$g(x) = \kappa x^{\lambda} e^{cx} \ \mathit{for}\ \kappa, \lambda, c \in \mathbb{R}$$

Proof

Since $\forall x \in (0,1)$ $g(x) > 0$

$$\frac{g(xy).g(x(1-y))}{g(\frac{y}{2}).g(\frac{1-y}{2})} = G(x) > 0$$

taking log's gives

$$\log(\frac{g(x(1-y))}{g(\frac{1-y}{2})}) = \log G(x) - \log(\frac{g(xy)}{g(\frac{y}{2})})$$

then differentiating with respect to x we obtain

$$(1-y)\frac{g'(x(1-y))}{g(x(1-y))} = \frac{\partial}{\partial x}\log G(x) - y.\frac{g'(xy)}{g(xy)} \quad [1]$$

Similarly differentiating with respect to y gives

$$-x\frac{g'(x(1-y))}{g(x(1-y))} + \frac{1}{2}\frac{g'(\frac{1-y}{2})}{g(\frac{1-y}{2})} = -x\frac{g'(xy)}{g(xy)} + \frac{1}{2}\frac{g'(\frac{y}{2})}{g(\frac{y}{2})} \quad [2]$$

then $x[1] + (1-y)[2]$ gives

$$x\frac{g'(xy)}{g(xy)} = x\frac{\partial}{\partial x}\log G(x) + \frac{1}{2}(1-y).\frac{g'(\frac{y}{2})}{g(\frac{y}{2})} - \frac{1}{2}(1-y).\frac{g'(\frac{1-y}{2})}{g(\frac{1-y}{2})}$$

$$= K(x) + C(y)$$

Note that since g is twice differentiable it follows that $K(x)$ is differentiable

Integrating with respect to y gives

$$g(xy) = U(x)V(y)e^{K(x).y}$$

Note again that $V(y)$ is differentiable since g is differentiable and U is differentiable since g and K are differentiable

$$\therefore U(x).V(y).e^{K(x).y} = U(y).V(x).e^{K(y).x}$$

$$\frac{V(y)}{U(y)} = \frac{V(x)}{U(x)}.e^{K(y).x-K(x).y}$$

Given the differentiability of K, U and V we differentiate with respect to x to obtain

$$e^{K(y).x-K(x).y}(\frac{\partial}{\partial x}\frac{V(x)}{U(x)} + \frac{V(x)}{U(x)}.(K(y) - K'(x).y))$$

$$-\frac{\frac{\partial}{\partial x}\frac{V(x)}{U(x)}}{\frac{V(x)}{U(x)}} = K(y) - K'(x).y$$

differentiating with respect to y gives

$$K'(y) - K'(x) = 0$$

$$\implies K(x) = ax + b$$

where a and b are arbitrary real constants

$$\therefore g(xy) = U(x)V(y)e^{axy}.e^{by}$$

Let $V(y).e^{by} = W(y)$ and note that W is differentiable since V is differentiable

This gives

$$g(xy) = W(y).U(x).e^{axy}$$

Then again since g(xy)=g(yx) we obtain

$$W(y).U(x) = W(x).U(y)$$

Differentiating with respect to x gives $W(x) = d.U(x)$ where d is an arbitrary real constant

$$\therefore g(xy) = d.U(x).U(y).e^{axy}$$

Making the change of variables to x, z where $z = x.y$ gives

$$g(z) = d.U(x).U(z/x).e^{a.z} \; , x \geq z$$

differentiating with respect to x

$$d.U'(x).U(\frac{z}{x}) - \frac{z}{x^2}.U'(\frac{z}{x})U(x),\ x \geq z$$

Making the change of variables to x, y where $y = \frac{z}{x}$ we obtain

$$d.U'(x).U(y) - \frac{y}{x}.U'(y).U(x)$$

$$\therefore\ y.\frac{U'(y)}{U(y)} = d.x.\frac{U'(x)}{U(x)}$$

$$\therefore\ \frac{U'(x)}{U(x)} = \frac{\lambda}{x}$$

where λ is an arbitrary real constant

Integrating with respect to x and making a trivial substitution gives the required result. $\Box$

Lemma 49 *Let the prior $f^{(n)} \in \mathcal{C}_M$ satisfy Strong Non Nullity, Weak Renaming and Relative Ignorance then $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$ for $g : (0,1) \to \mathbb{R}^{>0}$ and*

$$\forall x, y \in (0,1)\ g(xy)g(x(1-y)) \propto \frac{g * g(x)}{x}.g(\frac{y}{2}).g(\frac{1-y}{2})$$

Proof

By lemma 44 we have that $f^{(n)}$ is a symmetric multiplicative prior so that

$$\forall \vec{x} \in Int(V^{(n)}) f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$$

for $g : (0,1) \to \mathbb{R}^{>0}$

Then by Relative Ignorance

$\forall x \in (0,1)$ $\forall 0 < \delta < Min(x, 1-x)$ $\forall I$ an open interval of $(0,1)$ we have that

$$\frac{\int_{x-\delta}^{x+\delta} \int_{Is} g(t_1)g(s-t_1)dt_1 \int_0^{1-s} \ldots \int_0^{1-s-\sum_{i=3}^{2^n-2} t_i} \prod_{i=3}^{2^n} g(t_i)d\vec{t}ds}{\int_{x-\delta}^{x+\delta} \int_0^s g(t_1)g(s-t_1)dt_1 \int_0^{1-s} \ldots \int_0^{1-s-\sum_{i=3}^{2^n-2} t_i} \prod_{i=3}^{2^n} g(t_i)d\vec{t}ds} = H(I)$$

where H is some function on open intervals into $(0,1)$. Now since $f^{(n)} \in \mathcal{C}_M$ we have by proposition 35 that

$$\int_0^s g(t_1)g(s-t_1)dt_1 \int_0^{1-s} \ldots \int_0^{1-s-\sum_{i=3}^{2^n-2} t_i} \prod_{i=3}^{2^n} g(t_i)d\vec{t}$$

exists and is continuous at s $\forall s \in (0,1)$. Thus taking the limit as δ tends to 0 we have, given Strong Non Nullity, that $\forall x \in (0,1)$

$$\frac{\int_{xI} g(t)g(x-t)dt}{g * g(x)} = H(I)$$

Now making a change of variables to x, s where $t = xs$ gives

$$\frac{\int_I g(xs)g(x(1-s))xds}{g * g(x)} = H(I)$$

Now letting $x = \frac{1}{2}$ we have that

$$H(I) \propto \int_I g(\frac{s}{2})g(\frac{1-s}{2})ds$$

$$\therefore \frac{\int_I g(xs)g(x(1-s))xds}{\int_I g(\frac{s}{2})g(\frac{1-s}{2})ds} \propto \; g * g(x)$$

Since $f^{(n)} \in \mathcal{C}_M$ $\forall x \in (0,1]$ $\forall I$ an open subinterval of $(0,1)$ $\forall s \in I$ $g(xs)g(x(1-s))x$ is continuous at s. Hence, letting $I = (y-\delta, y+\delta)$ and taking the limit as δ tends to 0 we have the required result. $\square$

Proof of Theorem 47

($\Rightarrow$) Follows immediately from lemma 48, lemma 49

($\Leftarrow$)

Suppose $f^{(n)} = d(\lambda, n)$ for $\lambda \in \mathbb{R}^{>0}$ then

$$Prob^{(n)}(E(p_n|\alpha_i^{(n-1)}) < x|E(\alpha_i^{(n-1)} \in I)$$

$$= \frac{\int_I \int_0^{xt} s_1^{\frac{\lambda}{2^n}-1}(t-s_1)^{\frac{\lambda}{2^n}-1} \int_0^{1-t} \ldots \int_0^{1-t-\sum_{i=3}^{2^n-2} s_i} \prod_{i=3}^{2^n-1} s_i^{\frac{\lambda}{2^n}-1}(1-t-\sum_{i=3}^{2^n-1} s_i)^{\frac{\lambda}{2^n}-1} d\vec{s}dt}{\int_I \int_0^{t} s_1^{\frac{\lambda}{2^n}-1}(t-s_1)^{\frac{\lambda}{2^n}-1} \int_0^{1-t} \ldots \int_0^{1-t-\sum_{i=3}^{2^n-2} s_i} \prod_{i=3}^{2^n-1} s_i^{\frac{\lambda}{2^n}-1}(1-t-\sum_{i=3}^{2^n-1} s_i)^{\frac{\lambda}{2^n}-1} d\vec{s}dt}$$

Now for any integral

$$\int_0^{kt} s^a(t-s)^b ds$$

where $a, b > -1$ $k \in (0,1]$ we have, making the change of variables to y where $ty = s$, that

$$\int_0^{kt} s^a(t-s)^b ds = t^{a+b-1} \int_0^k y^a(1-y)^b dy$$

$\therefore$

$$Prob^{(n)}(E(p_n|\alpha_i^{(n-1)}) < x|E(\alpha_i^{(n-1)} \in I)$$
$$= \frac{\int_0^x y^{\frac{\lambda}{2^n}-1}(1-y)^{\frac{\lambda}{2^n}-1} dy \int_I t^{\frac{\lambda}{2^n}-1}(1-t)^{(2^n-3)\frac{\lambda}{2^n}-1} dt}{\int_0^1 y^{\frac{\lambda}{2^n}-1}(1-y)^{\frac{\lambda}{2^n}-1} dy \int_I t^{\frac{\lambda}{2^n}-1}(1-t)^{(2^n-3)\frac{\lambda}{2^n}-1} dt}$$

$$= \frac{\int_0^x y^{\frac{\lambda}{2^n}-1}(1-y)^{\frac{\lambda}{2^n}-1}dy}{\int_0^1 y^{\frac{\lambda}{2^n}-1}(1-y)^{\frac{\lambda}{2^n}-1}dy}$$

□

The following corollary shows that for the class of densities $\mathcal{C}_M^2$, the axioms A2, A3, A4 and A7 characterise the hierarchy of symmetric Dirichlet priors up to a parameter λ independent of n.

Corollary 50 *If $\{f^{(n)}\}$ is a hierarchy of priors satisfying Marginality such that for all $n \geq 1$ $f^{(n)}$ is in $\mathcal{C}_M^2$ and satisfies Non Nullity, Weak Renaming and Relative Ignorance, then $f^{(n)}(\vec{x}) = d(\lambda, 2^n)(\vec{x})$ where $\lambda > 0$ is a constant independent of n.*

Proof

Given Non Nullity and Marginality we have by lemma 46 that $\forall n > 0$ $f^{(n)}$ satisfies Strong Non Nullity. Then by Weak Renaming, Relative Ignorance and theorem 47 we have that $\forall n > 0$ $f^{(n)}$ is a symmetric Dirichet prior. Further given Marginality we have by proposition 25 that $\forall n > 0$ $f^{(n)} = d(\lambda, n)$ for $\lambda > 0$ independent of n □

2.4 Comments

It is important to note that we are not claiming the above axioms are natural laws justifying Dirichlet priors. Rather the aim is to demonstrate the relationship between particular hierarchies of priors and epistemological conditions. Indeed, in the next chapter we discuss a number of other axioms some of which, for hierarchies of Dirichlet priors, are inconsistent with principles described above.

An interesting open problem is whether the assumption that $f^{(n)} \in \mathcal{C}^2_M$ in the above can be replaced by $f^{(n)} \in \mathcal{C}^2$. This would seem, in part, to be dependent on the following conjecture. If $f^{(n)} \in \mathcal{C}$ is a symmetric multiplicative prior then there is some function $g : [0,1] \to \mathbb{R}^{\geq 0}$ continuous on $(0,1)$ such that $\forall \vec{x} \in V^{(n)}\ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$.

Chapter 3

Axioms of Conditionality and Improbability

We now introduce a number of principles relating to the evaluation of conditional probability values. In addition, we propose an axiom concerning the problem of how a reasoning process should respond when a new propositional variable p_n is added to the language about which nothing is known except that $E(p_n)$ is arbitrarily small.

3.1 Axioms of Conditionality

A8 Conditional Ignorance

$\forall n > 0$ $Prob^{(n)}$ satisfies Weak Renaming and

$$Prob^{(n)}(E(p_j|\bigwedge_{i\neq j} p_i) \in I) \;=\; Prob^{(1)}(E(p_j) \in I)$$

for I a strict interval of $[0,1]$

Given that the events $\underline{p_j} = 1$ for $1 \leq i \leq 2^n$ $i \neq j$ have occurred the 2 events $\underline{\bigwedge_{i\neq j} p_i \wedge p_j} = 1$ and $\underline{\bigwedge_{i\neq j} p_i \wedge \neg p_j} = 1$ are mutually exclusive and determine all the possible outcomes of the random propositional variables of $L^{(n)}$. Hence, the problem of calculating the probability of $\underline{p_j} = 1$ when all that is known is that events $\underline{p_i} = 1$ for $1 \leq i \leq 2^n$ $i \neq j$ have occurred should be equivalent to the problem of calculating the probability of $\underline{p_j} = 1$, *a priori*, when p_j is the only propositional variable in the language.

A9 Strong Conditional Ignorance

$\forall n > 0$ $Prob^{(n)}$ satisfies Weak Renaming and

$$Prob^{(n)}(E(p_j|\bigwedge_{i\neq j} p_i) \in J \mid E(\bigwedge_{i\neq j} p_i) \in I) = Prob^{(1)}(E(p_j) \in J)$$

for J and I strict intervals of $[0,1]$

This is a strengthening of Conditional Ignorance which informally states that the problem of calculating the probability that $\underline{p_j} = 1$ when the only knowledge is that the events $\underline{p_i} = 1$ for $1 \leq i \leq 2^n$ $i \neq j$ have occurred and that the probability of these events occurring together is in some strict interval of $[0,1]$ is equivalent to the problem of calculating the probability of $\underline{p_i} = 1$, *a priori*, when p_j is the only propositional variable under consideration.

3.2 An Axiom of Improbability

Notation: For $\vec{x}, \vec{y} \in \mathbb{R}^n$ we write $\vec{x} < \vec{y}$ iff $x_i < y_i$ for $i = 1, \ldots, n$

A10 Total Improbability

Let $\{f^{(n)}\}$ be a hierarchy of priors such that $\forall n > 0\ f^{(n)}$ satisfies Weak Renaming. Then given the conditionalisation assumption $\{f^{(n)}\}$ satisfies Total Improbability iff

$$\forall \delta > 0\ \exists \vec{\Delta} \in \mathbb{R}^{2^{n-1}}\ \vec{\Delta} > \vec{0}\ \forall \vec{0} < \vec{\epsilon} < \vec{\Delta}$$

$$|Prob^{(n)}(\bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)}) \in I_i | \bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)} \wedge p_n) = \epsilon_i) - Prob^{(n-1)}(\bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)}) \in I_i)| < \delta$$

where I_i is a strict interval of $[0,1]$ for $i = 1, \ldots, 2^n$

As an abbreviation we write

$$\lim_{\vec{\epsilon} \to \vec{0}} Prob^{(n)}(\bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)}) \in I_i | \bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)} \wedge p_n) = \epsilon_i) =$$

$$Prob^{(n-1)}(\bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)}) \in I_i)$$

For any atom $\alpha_i^{(n-1)}$ the problem of giving a belief value to $\alpha_i^{(n-1)} \wedge \neg p_n$ while knowing that $E(\alpha_j^{(n-1)} \wedge p_n)$ is arbitrarily small for all $\alpha_j^{(n-1)}$ $j = 1, \ldots, 2^{n-1}$ is equivalent to the problem of giving an *a priori* belief value to $\alpha_i^{(n-1)}$ when p_n is not in the language.

The idea is that the expert only knows that the situations when $\alpha_i^{(n-1)} \wedge \neg p_n$ occurs are exactly the situations when $\alpha_i^{(n-1)}$ occurs but he has no information about these situations.

The next result illustrates the relationship between Total Improbability and hierarchies of symmetric multiplicative priors for which $g^{(n)} \propto g^{(n-1)}$

Theorem 51 *If* $\forall n > 0$ $f^{(n)} \in \mathcal{C}_M$ *and satisfies Weak Renaming and Strong Non Nullity then given the conditionalisation assumption* $\{f^{(n)}\}$ *satisfies Total Improbability iff* $\forall n > 0$ $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g^{(n)}(x_i)$ *where* $g^{(n)} : [0,1] \to \mathbb{R}^{\geq 0}$ *is such that* $\prod_{i=1}^{2^{n-1}} g^{(n)}(x_i)$ *is integrable on* $V^{(n-1)}$ *and*

$$\forall \vec{x} \in Int(V^{(n-1)})\ f^{(n-1)}(\vec{x}) = \frac{\prod_{i=1}^{2^{n-1}} g^{(n)}(x_i)}{\int_{V^{(n-1)}} \prod_{i=1}^{2^{n-1}} g^{(n)}(x_i) dV^{(n-1)}}$$

Proof

($\Rightarrow$)

Given Weak Renaming and Strong Non Nullity we have by lemma 44 that $\forall n > 0$ $f^{(n)}$ is a symmetric multiplicative prior. In other words

$$\forall \vec{x} \in Int(V^{(n)}) f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g^{(n)}(x_i)$$

By Total Improbability $\forall \vec{y} \in Int(V^{(n-1)})$ $\forall \delta > 0$

$$\lim_{\vec{\epsilon} \to \vec{0}} Prob^{(n)}(\bigwedge_{i=1}^{2^{n-1}-1} E(\alpha_i^{(n-1)}) \in (y_i - \delta, y_i + \delta) \mid \bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)} \wedge p_n) = \epsilon_i)$$
$$= Prob^{(n-1)}(\bigwedge_{i=1}^{2^{n-1}-1} E(\alpha_i^{(n-1)}) \in (y_i - \delta, y_i + \delta))$$

Now making a change of variables to t_i, x_{2i-1} where $x_{2i-1} + x_{2i} = t_i$ for $i = 1, \ldots, 2^{n-1}$ we have given Strong Non Nullity for δ sufficiently small that

$$\lim_{\vec{\epsilon} \to \vec{0}} \frac{\int_{y_1-\delta}^{y_1+\delta} \cdots \int_{y_{2^{n-1}-1}-\delta}^{y^{2^{n-1}-1}+\delta} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i - \epsilon_i) d\vec{t}}{\int_{\epsilon_1}^{1-\sum_{i=2}^{2^{n-1}} \epsilon_i} \cdots \int_{\epsilon_{2^{n-1}-1}}^{1-\epsilon_{2^{n-1}}-\sum_{j=1}^{2^{n-1}-2} t_j} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i - \epsilon_i) d\vec{t}}$$

$$= \int_{y_1-\delta}^{y_1+\delta} \cdots \int_{y_{2^{n-1}-1}-\delta}^{y_{2^{n-1}-1}+\delta} \prod_{i=1}^{2^{n-1}} g^{(n-1)}(t_i) d\vec{t} \qquad [1]$$

Now since $g^{(n)}$ is uniformly continuous on $[y_i - \Delta, y_i + \Delta]$ $\forall y_i \in (0,1)$ and $\forall \Delta$ such that $0 < \Delta < min(y_i, 1-y_i)$ we have that

$$\lim_{\vec{\epsilon}\to\vec{0}} \int_{y_1-\delta}^{y_1+\delta} \cdots \int_{y_{2^{n-1}-1}-\delta}^{y_{2^{n-1}-1}+\delta} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i - \epsilon_i) d\vec{t} = \int_{y_1-\delta}^{y_1+\delta} \cdots \int_{y_{2^{n-1}-1}-\delta}^{y_{2^{n-1}-1}+\delta} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i) d\vec{t}$$

Furthermore, making the change of variables to $s_i = t_i - \epsilon_i$ for $i = 1, \ldots, 2^{n-1} - 1$ we have that

$$\int_{\epsilon_1}^{1-\sum_{i=2}^{2^{n-1}} \epsilon_i} \cdots \int_{\epsilon_{2^{n-1}-1}}^{1-\epsilon_{2^{n-1}}-\sum_{j=1}^{2^{n-1}-2} t_j} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i - \epsilon_i) d\vec{t}$$

$$= \int_0^{1-\sum_{i=1}^{2^{n-1}} \epsilon_i} \cdots \int_0^{1-\sum_{i=1}^{2^{n-1}} \epsilon_i - \sum_{i=1}^{2^{n-1}-2} s_i} \prod_{i=1}^{2^{n-1}} g^{(n)}(s_i) d\vec{s} = g^{(n)\ *\ 2^{n-1}}(1 - \sum_{i=1}^{2^{n-1}} \epsilon_i)$$

Notice, that if $\prod_{i=1}^{2^{n-1}} g^{(n)}(x_i)$ is integrable on $V^{(n-1)}$ then

$$\int_{V^{(n-1)}} \prod_{i=1}^{2^{n-1}} g^{(n)}(x_i) dV^{(n-1)} = g^{(n)\ *\ 2^{n-1}}(1)$$

Now suppose that $\prod_{i=1}^{2^{n-1}} g^{(n)}(x_i)$ is not integrable on $V^{(n-1)}$ then $g^{(n)\ *\ 2^{n-1}}(1)$ does not exist. Therefore, by proposition 35 $g^{(n)\ *\ 2^{n-1}}(x)$ is unbounded as x tends to 1. Therefore, $\forall \vec{y} \in Int(V^{(n-1)})$ and $\forall \delta$ such that $0 < \delta < min(y_i, 1-y_i)$ for $i = 1, \ldots, 2^{n-1}$

$$\lim_{\vec{\epsilon}\to\vec{0}} \frac{\int_{y_1-\delta}^{y_1+\delta} \cdots \int_{y_{2^{n-1}-1}-\delta}^{y_{2^{n-1}-1}+\delta} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i - \epsilon_i) d\vec{t}}{g^{(n)\ *\ 2^{n-1}}(1 - \sum_{i=1}^{2^{n-1}} \epsilon_i)} = 0$$

Hence,

$$\int_{y_1-\delta}^{y_1+\delta} \cdots \int_{y_{2^{n-1}-1}-\delta}^{y_{2^{n-1}-1}+\delta} \prod_{i=1}^{2^{n-1}} g^{(n-1)}(t_i) d\vec{t} = 0$$

$$\Rightarrow \ \forall \vec{y} \in Int(V^{(n-1)}) \ \prod_{i=1}^{2^{n-1}} g^{(n-1)}(y_i) = 0$$

which is a contradiction since $\prod_{i=1}^{2^{n-1}} g^{(n-1)}(y_i)$ is a density function on $V^{(n-1)}$

Therefore,

$$L.H.S \ of \ [1] \ = \frac{\int_{y_1-\delta}^{y_1+\delta} \cdots \int_{y^{2^{n-1}-1}-\delta}^{y^{2^{n-1}-1}+\delta} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i) d\vec{t}}{\int_0^1 \cdots \int_0^{1-\sum_{j=1}^{2^{n-1}-2} t_j} \prod_{i=1}^{2^{n-1}} g^{(n)}(t_i) d\vec{t}}$$

and then letting $\delta \to 0$ gives by continuity of $g^{(n)}$ and $g^{(n-1)}$ on $(0,1)$ and Strong Non Nullity the required result

($\Leftarrow$)

Consider

$$\lim_{\vec{\epsilon}\to\vec{0}} Prob^{(n)}(\bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)}) \in [a_i, b_i] | \bigwedge_{i=1}^{2^{n-1}} E(\alpha_i^{(n-1)} \wedge p_n) = \epsilon_i)$$

where $0 \leq a_i < b_i \leq 1$ for $i = 1, \ldots, 2^{n-1}$

After cancelling $\prod_{i=1}^{2^{n-1}} g(\epsilon_i)$ and making a change of variables we have that this is equal to

$$\lim_{\vec{\epsilon}\to\vec{0}} \frac{\int_{v_1(\epsilon)}^{u_1(\vec{\epsilon})} \cdots \int_{v_{2^{n-1}-1}(\vec{\epsilon},s_1,\ldots,s_{2^{n-1}-2})}^{u_{2^{n-1}-1}(\vec{\epsilon},s_1,\ldots,s_{2^{n-1}-2})} \prod_{i=1}^{2^{n-1}} g^{(n)}(s_i)d\vec{s}}{g^{(n)\ *2^{n-1}}(1-\sum_{i=1}^{2^{n-1}} \epsilon_i)} \qquad [1]$$

where

$$u_i(\vec{\epsilon}, s_1, \ldots, s_{i-1}) = min(b_i - \epsilon_i, 1 - \sum_{j=1}^{2^{n-1}-1} \epsilon_i - a_{2^{n-1}} - \sum_{j=1}^{i-1} s_j, 1 - \sum_{j=1}^{2^{n-1}} \epsilon_i - \sum_{j=1}^{i-1} s_j)$$

and

$$v_i(\vec{\epsilon}, s_1, \ldots, s_{i-1}) = max(a_i - \epsilon_i, 1 - b_i - \sum_{j=1}^{2^{n-1}-1} \epsilon_i - \sum_{j=1}^{i-1} s_j, 0)$$

Now since $g^{(n)}$ is continuous on $(0,1)$ we have that the denominator in [1] is a continuous function of $\vec{\epsilon}$ on $ISEG^{(2^{n-1})}(1)$. Also, since $g^{(n)\ *2^{n-1}}(1)$ exists it follows that the denominator in [1] is bounded as $\vec{\epsilon}$ tends to $\vec{0}$.

$$\therefore\ [1] = \frac{\int_{v_1'}^{u_1'} \cdots \int_{v'_{2^{n-1}-1}(s_1,\ldots,s_{2^{n-1}-2})}^{u'_{2^{n-1}-1}(s_1,\ldots,s_{2^{n-1}-2})} \prod_{i=1}^{2^{n-1}} g(s_i)d\vec{s}}{g^{(n)\ *2^{n-1}}(1)}$$

where

$$u_i'(s_1, \ldots, s_{i-1}) = min(b_i, 1 - a_{2^{n-1}} - \sum_{j=1}^{i-1} s_j)$$

and

$$v_i'(s_1, \ldots, s_{i-1}) = max(a_i, 1 - b_{2^{n-1}} - \sum_{j=1}^{i-1} s_j)$$

$$= Prob^{(n-1)}(\bigwedge_{j=1}^{2^{n-1}} E(\alpha_j^{(n-1)}) \in [a_j, b_j])$$

□

Corollary 52 *If* $\forall n \geq 1$ $f^{(n)} \in \mathcal{C}_M^2$ *and satisfies Strong Non Nullity, Weak Renaming and Relative Ignorance then given the conditionalisation assumption the following are equivalent* *(i)* $\{f^{(n)}\}$ *satisfies Total Improbability*

(ii) $\{f^{(n)}\} = \{d(2^{n-1}\lambda, n)\}$ *for* $\lambda \in \mathbb{R}^{>0}$

Proof

$((i) \Rightarrow (ii))$

By Strong Non Nullity, Weak Renaming and Relative Ignorance for priors in $\mathcal{C}^2_M$ we have by theorem 47

$$\forall \vec{x} \in Int(V^{(n)}) \ f^{(n)}(\vec{x}) = d(\lambda_n, n)(\vec{x}) = \Gamma(\lambda_n)[\Gamma(\frac{\lambda_n}{2^n})]^{-2^n} \prod_{i=1}^{2^n} x_i^{\frac{\lambda_n}{2^n}-1}$$

Now by Total Improbability and theorem 51 we have

$$\prod_{i=1}^{2^{n-1}} g^{(n-1)}(x_i) = \frac{\prod_{i=1}^{2^{n-1}} g^{(n)}(x_i)}{\int_{V^{(n-1)}} \prod_{i=1}^{2^{n-1}} g^{(n)}(x_i) dV^{(n-1)}}$$

then

$$\Gamma(\lambda_{n-1})[\Gamma(\frac{\lambda_{n-1}}{2^{n-1}})]^{-2^{n-1}} \prod_{i=1}^{2^{n-1}} x_i^{\frac{\lambda_{n-1}}{2^{n-1}}-1} = \frac{\prod_{i=1}^{2^{n-1}} x_i^{\frac{\lambda_n}{2^n}-1}}{\int_{V^{(n-1)}} \prod_{i=1}^{2^{n-1}} x_i^{\frac{\lambda_n}{2^n}-1} dV^{(n-1)}}$$

$$\prod_{i=1}^{2^{n-1}} x_i^{\frac{\lambda_{n-1}}{2^{n-1}}-1} = \kappa_n \prod_{i=1}^{2^{n-1}} x_i^{\frac{\lambda_n}{2^n}-1}$$

for $\kappa_n \in \mathbb{R}^{>0}$. Since functions of the form x^β $\beta > -1$ are continuous and strictly positive on $(0,1)$ then by proposition 27

$$x^{\frac{\lambda_{n-1}}{2^{n-1}}-1} = a_n e^{b_n x} x^{\frac{\lambda_n}{2^n}-1}$$

$$\Rightarrow \ b_n = 0 \ \Rightarrow \frac{\lambda_{n-1}}{2^{n-1}} = \frac{\lambda_n}{2^n}$$

$((ii) \Rightarrow (i))$

$\forall n > 0 \ n \in \mathbb{N}$ suppose that

$$\forall \vec{x} \in Int(V^{(n-1)}) \ f^{(n-1)}(\vec{x}) = d(\lambda 2^{n-1}, n)(\vec{x})$$

$$= \frac{\prod_{i=1}^{2^{n-1}} x_i^{\frac{\lambda}{2}-1}}{\int_{V^{(n-1)}} \prod_{i=1}^{2^{n-1}} x_i^{\frac{\lambda}{2}-1} d\vec{x}} = \frac{d(\lambda 2^n, n)(\vec{x})}{\int_{V^{(n-1)}} d(\lambda 2^n, n)(\vec{x}) d\vec{x}}$$

□

3.3 Total Improbability and Conditional Ignorance

In this section we prove a result showing that for hierarchies of priors in $\mathcal{C}_M^2$ satisfying Strong Non Nullity, Weak Renaming, and Relative Ignorance, the principles of Conditional Ignorance and Total Improbability are equivalent.

Lemma 53 *If for all $n > 0$ $f^{(n)} \in \mathcal{C}_M$ and satisfies Strong Non Nullity and Weak Renaming then the following are equivalent*

(i) $\forall n > 0$ $f^{(n)}$ satisfies Relative Ignorance and $\{f^{(n)}\}$ satisfies Conditional Ignorance

(ii) $\{f^{(n)}\}$ satisfies Strong Conditional Ignorance

Proof

$((i) \Rightarrow (ii))$

For $Prob^{(n)}$ satisfying Relative Ignorance

$$Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in J \mid E(\alpha_1^{(n-1)}) \in I) = Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in J)$$

where I and J are strict intervals of $[0,1]$

$$= Prob^{(1)}(E(p_n) \in J)$$

by Conditional Ignorance

Then given Weak Renaming we have Strong Conditional Ignorance

$((ii) \Rightarrow (i))$

$$Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in J \mid E(\alpha_1^{(n-1)}) \in I) = Prob^{(1)}(E(p_n) \in J)$$

$\Rightarrow Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in J \mid E(\alpha_1^{(n-1)}) \in I)$ is independent of I

$\Rightarrow$ given Weak Renaming $Prob^{(n)}$ satisfies Relative Ignorance

and as above we have

$$Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in J \mid E(\alpha_1^{(n-1)}) \in I) = Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in J)$$

$$\therefore Prob^{(n)}(E(p_n|\alpha_1^{(n-1)}) \in J) = Prob^{(1)}(E(p_n) \in J)$$

Then in the presence of Weak Renaming we have the required result. $\Box$

Lemma 54 *If* $\forall n > 0$ $f^{(n)} \in \mathcal{C}_M^2$ *and satisfies Strong Non Nullity and Weak Renaming then* $\{f^{(n)}\}$ *satisfies Strong Conditional Ignorance iff* $\{f^{(n)}\} = \{d(2^{n-1}\lambda, n)\}$ *for* $\lambda \in \mathbb{R}^{>0}$

Proof

$(\Rightarrow)$

Given Strong Non Nullity and Strong Conditional Ignorance then by lemma 53 $\forall n > 0$ $f^{(n)}$ satisfies Relative Ignorance

Now by Relative Ignorance and Weak Renaming on $f^{(n)} \in \mathcal{C}^2_M$ and theorem 47

$$\forall n \geq 1 \; f^{(n)}(\vec{x}) = d(\lambda_n, n)(\vec{x}) \; \lambda_n > 0$$

Then by Conditional Ignorance

$$\frac{\int_0^1 \int_{It} x^{\frac{\lambda_n}{2^n}-1}(t-x)^{\frac{\lambda_n}{2^n}-1}dx(1-t)^{(2^n-2)\frac{\lambda_n}{2^n}-1}dt}{\int_0^1 \int_0^t x^{\frac{\lambda_n}{2^n}-1}(t-x)^{\frac{\lambda_n}{2^n}-1}dx(1-t)^{(2^n-2)}} = \int_I x^{\frac{\lambda}{2}-1}(1-x)^{\frac{\lambda}{2}-1}dx$$

where I is a strict interval of $[0,1]$ $t \in (0,1)$ and $\lambda > 0$ is independent of n

Then making a change of variables to t, s where $s = tx$ gives that

$$\frac{\int_I x^{\frac{\lambda_n}{2^n}-1}(1-x)^{\frac{\lambda_n}{2^n}-1}dx}{\int_0^1 x^{\frac{\lambda_n}{2^n}-1}(1-x)^{\frac{\lambda_n}{2^n}-1}dx} = \int_I x^{\frac{\lambda}{2}-1}(1-x)^{\frac{\lambda}{2}-1}dx$$

Then letting $I = (y-\delta, y+\delta)$ and $\delta \to 0$ gives

$$y^{\frac{\lambda_n}{2^n}-1}(1-y)^{\frac{\lambda_n}{2^n}-1} = \kappa_n y^{\frac{\lambda}{2}-1}(1-y)^{\frac{\lambda}{2}-1}$$

for $\kappa_n \in \mathbb{R}^{>0}$

$$y^{\frac{\lambda_n}{2^n}-\frac{\lambda}{2}}(1-y)^{\frac{\lambda_n}{2^n}-\frac{\lambda}{2}} = \kappa_n$$
$$\Rightarrow \frac{\lambda_n}{2^n} = \frac{\lambda}{2}$$

$(\Leftarrow)$

By theorem 47 $\{d(2^{n-1}\lambda, n)\}$ satisfies Relative Ignorance and for I a strict interval of $[0,1]$

$$Prob^{(n)}(E(p_j|\bigwedge_{i\neq j} p_i) \in I) = Prob^{(n)}(E(p_n|\bigwedge_{i=1}^{n-1} p_i) \in I)$$
$$= \frac{\int_0^1 \int_{It} x^{\frac{\lambda}{2}-1}(t-x)^{\frac{\lambda}{2}-1}dx(1-t)^{(2^n-2)\frac{\lambda}{2}-1}dt}{\int_0^1 \int_0^t x^{\frac{\lambda}{2}-1}(t-x)^{\frac{\lambda}{2}-1}(1-t)^{(2^n-2)\frac{\lambda}{2}-1}dt}$$

Now making a change of variables to s, t where $s = xt$ gives

$$\frac{\int_I s^{\frac{\lambda}{2}-1}(1-s)^{\frac{\lambda}{2}-1}ds}{\int_0^1 s^{\frac{\lambda}{2}-1}(1-s)^{\frac{\lambda}{2}-1}ds} = Prob^{(1)}(E(p_n) \in I)$$

□

Theorem 55 *If* $\forall n \geq 1\ f^{(n)} \in \mathcal{C}_M^2$ *and satisfies Strong Non Nullity, Weak Renaming and Relative Ignorance then given the conditionalisation assumption the following are equivalent*

(i) $\{f^{(n)}\}$ *satisfies Conditional Ignorance*

(ii) $\{f^{(n)}\}$ *satisfies Total Improbability*

Proof

$((i) \Rightarrow (ii))$

$\forall n > 0\ f^{(n)} \in \mathcal{C}_M^2$ and satisfies Strong Non Nullity, Weak Renaming and Relative Ignorance and $\{f^{(n)}\}$ satisfies Conditional Ignorance then by lemma 53 $\{f^{(n)}\}$ satisfies Strong Conditional Ignorance $\Rightarrow$ by lemma 54 that $\{f^{(n)}\} = \{d(2^{n-1}\lambda, n)\}$ for $\lambda \in \mathbb{R}^{>0}$ $\Rightarrow$ by corollary 52 that $\{f^{(n)}\}$ satisfies Total Improbability

$((ii) \Rightarrow (i))$

If $\forall n \geq 1\ f^{(n)} \in \mathcal{C}_M^2$ and satisfies Strong Non Nullity, Weak Renaming and Relative Ignorance and $\{f^{(n)}\}$ satisfies Total Improbability then by corollary 52 $\{f^{(n)}\} = \{d(2^{n-1}\lambda, n)\}$ $\Rightarrow$ by lemma 54 that $\{f^{(n)}\}$ satisfies Strong Conditional Ignorance $\Rightarrow$ by lemma 53 $\{f^{(n)}\}$ satisfies Conditional Ignorance $\square$

In other words, for hierarchies of suitably smooth symmetric multiplicative priors, in the presence of Strong Non Nullity and Relative Ignorance, the principles of Conditional Ignorance and Total Improbability are equivalent. This seems some what surprising given the different epistemological notions underlying these two principles.

In view of the justification for Conditional Ignorance you might consider strengthening the axiom so that $\forall S \subseteq \{1, \ldots, n\} - \{i\}$

$$Prob^{(n)}(E(p_i | \bigwedge_{j \in S} p_j) \in I) = Prob^{(n-k)}(E(p_i) \in I)$$

where $k = |S|$ and I is a strict interval of $[0, 1]$

In other words, given that the events $\underline{p_j} = 1$ for $j \in S$ have occurred then the 2^{n-k} events $\underline{\bigwedge_{j \in S} p_j \wedge \beta_i} = 1$, for $i = 1, \ldots, 2^{n-k}$ where $\{\beta_j\}$ is an enumeration of the atoms of $L^{(n)} - \{p_j | \ j \in S\}$, are mutually exclusive and determine all the possible outcomes of the random propositional variables of $L^{(n)}$. Hence, the problem of calculating the probability of the event $\underline{p_i} = 1$ when all that is known is that the events $\underline{p_j} = 1$ for $j \in S$ have occurred should be equivalent to the problem of calculating the probability of $\underline{p_i} = 1$, *a priori* , when only the propositional variables p_j for $j \notin S$ are considered.

Proposition 56 *If $\{f^{(n)}\}$ is a hierarchy of symmetric Dirichlet priors then the following are equivalent*

(i) $\{f^{(n)}\}$ satisfies $\forall S \subseteq \{1, \ldots n\} - \{i\}$

$$Prob^{(n)}(E(p_i | \bigwedge_{j \in S} p_j) \in I) = Prob^{(n-k)}(E(p_i) \in I)$$

where $k = |S|$ and I is a strict interval of $[0, 1]$

(ii) $\{f^{(n)}\} = \{d(2^{n-1}\lambda, n)\}$ for $\lambda \in \mathbb{R}^{>0}$

Proof

$((i) \Rightarrow (ii))$

Let $S = \{1, \dots n\} - \{i\}$ then the result follows immediately by lemma 54

$((ii) \Rightarrow (i))$

Let $S \subseteq \{1, \dots, n\} - \{i\}$ such that $|S| = k$ for $k \in \{1, \dots, n-1\}$

w.l.o.g consider $Prob^{(n)}(E(p_n| \bigwedge_{i=1}^{k} p_i) < y)$

Let $t = 2^{n-k}$ and make the change of variables $s = \sum_{i=1}^{t} x_i$ and $r = \sum_{i=1}^{\frac{t}{2}} x_{2i-1}$. Hence, we obtain for I a strict interval of $[0, 1]$

$$Prob^{(n)}(E(p_n| \bigwedge_{i=1}^{k} p_i) \in I) = Prob^{(n)}(E(p_{k+1}| \bigwedge_{i=1}^{k} p_i) \in I)$$

$$\int_0^1 \int_{sI} \int_0^r \dots \int_0^{r-\sum_{i=1}^{\frac{t}{2}-2} x_{2i-1}} \int_0^{s-r} \dots \int_0^{s-r-\sum_{i=1}^{\frac{t}{2}-2} x_{2i}} \int_0^{1-s} \dots$$

$$\int_0^{1-s-\sum_{i=t+1}^{2^n-2} x_i} d(\lambda_n, n)(\vec{x}) dx_{2^n-1} \dots dx_{t+1} dx_{t-2}, \dots, dx_4 dx_2 dx_{t-3}, \dots, dx_3 dx_1 dr ds$$

where $\lambda_n = 2^{n-1}\lambda$ for $\lambda \in \mathbb{R}^{>0}$

$$= \Gamma(\lambda_n)[\Gamma(\frac{\lambda_n}{2^n})]^{-2^n} \int_0^1 \int_{sI} \int_0^r \dots \int_0^{r-\sum_{i=1}^{\frac{t}{2}-2} x_{2i-1}} \prod_{i=1}^{\frac{t}{2}-1} x_{2i-1}^{\frac{\lambda_n}{2^n}-1} (t - \sum_{i=1}^{\frac{t}{2}-1} x_{2i-1})^{\frac{\lambda_n}{2^n}-1}$$

$$\int_0^{s-r} \dots \int_0^{s-r-\sum_{i=1}^{\frac{t}{2}-2} x_{2i}} \prod_{i=1}^{\frac{t}{2}-1} x_{2i}^{\frac{\lambda_n}{2^n}-1} (s - r - \sum_{i=1}^{\frac{t}{2}-1} x_{2i})^{\frac{\lambda_n}{2^n}-1}$$

$$\int_0^{1-s} \dots \int_0^{1-s-\sum_{i=t+1}^{\frac{t}{2}-1}} \prod_{i=t+1}^{2^n-2} x_i^{\frac{\lambda_n}{2^n}-1} dx_{2^n-1} \dots dx_{t+1} dx_{t-2}, \dots, dx_4 dx_2 dx_{t-3}, \dots, dx_3 dx_1 dr ds$$

Now consider

$$\int_0^{1-s} \dots \int_0^{1-s-\sum_{i=t+1}^{2^n-2} x_i} \prod_{i=1}^{\frac{t}{2}-1} x_{2i}^{\frac{\lambda_n}{2^n}-1} dx_{2^n-1} \dots dx_{t+1}$$

$$= (1-s)^{\frac{(2^n-t)\lambda_n}{2^n}-1} \prod_{j=1}^{2^n-t-1} \beta(\frac{\lambda_n}{2^n}, j\frac{\lambda_n}{2^n})$$

also

$$\int_0^{s-r} \dots \int_0^{s-r-\sum_{i=1}^{\frac{t}{2}-2}} \prod_{i=1}^{\frac{t}{2}-1} x_{2i}^{\frac{\lambda_n}{2^n}-1} (s-r-\sum_{i=1}^{\frac{t}{2}-1} x_{2i})^{\frac{\lambda_n}{2^n}-1} dx_{t-2}dx_{t-4}\dots dx_4 dx_2$$

$$= (s-r)^{\frac{t}{2}\frac{\lambda_n}{2^n}-1} \prod_{j=1}^{\frac{t}{2}-1} \beta(\frac{\lambda_n}{2^n}, j\frac{\lambda_n}{2^n})$$

and

$$\int_0^{r} \dots \int_0^{r-\sum_{i=1}^{\frac{t}{2}-2} x_{2i-1}} \prod_{i=1}^{\frac{t}{2}-1} x_{2i-1}^{\frac{\lambda_n}{2^n}-1} (t-\sum_{i=1}^{\frac{t}{2}-1} x_{2i-1})^{\frac{\lambda_n}{2^n}-1} dx_{t-3}\dots dx_1$$

$$= r^{\frac{t}{2}\frac{\lambda_n}{2^n}-1} \prod_{j=1}^{\frac{t}{2}-1} \beta(\frac{\lambda_n}{2^n}, j\frac{\lambda_n}{2^n})$$

$\therefore$ we have $Prob^{(n)}(E(p_n| \wedge_{i=1}^k p_i) \in I) =$

$$\Gamma(\lambda_n)[\Gamma(\frac{\lambda_n}{2^n})]^{-2^n} (\prod_{j=1}^{\frac{t}{2}-1} \beta(\frac{\lambda_n}{2^n}, j\frac{\lambda_n}{2^n}))^2 \prod_{j=1}^{2^n-t-1} \beta(\frac{\lambda_n}{2^n}, j\frac{\lambda_n}{2^n})$$

$$\times \int_0^1 \int_{sI} r^{\frac{t}{2}\frac{\lambda_n}{2^n}-1} (s-r)^{\frac{t}{2}\frac{\lambda_n}{2^n}-1} (1-s)^{\frac{(2^n-t)\lambda_n}{2^n}-1} dr ds$$

$$= \int_0^1 (1-s)^{\frac{(2^n-t)\lambda_n}{2^n}-1} s^{t\frac{\lambda_n}{2^n}-1} ds \int_I r^{\frac{t}{2}\frac{\lambda_n}{2^n}-1} (1-r)^{\frac{t}{2}\frac{\lambda_n}{2^n}-1} dr$$

$$\times\Gamma(\lambda_n)[\Gamma(\frac{\lambda_n}{2^n})]^{-2^n}(\prod_{j=1}^{\frac{t}{2}-1}\beta(\frac{\lambda_n}{2^n},j\frac{\lambda_n}{2^n}))^2\prod_{j=1}^{2^n-t-1}\beta(\frac{\lambda_n}{2^n},j\frac{\lambda_n}{2^n})$$

$$\int_I r^{2^{-(k+1)}\lambda_n-1}(1-r)^{2^{-(k+1)}\lambda_n-1}dr$$

$$\times\beta(\frac{(2^n-t)\lambda_n}{2^n},\frac{t\lambda_n}{2^n})\Gamma(\lambda_n)[\Gamma(\frac{\lambda_n}{2^n})]^{-2^n}(\prod_{j=1}^{\frac{t}{2}-1}\beta(\frac{\lambda_n}{2^n},j\frac{\lambda_n}{2^n}))^2\prod_{j=1}^{2^n-t-1}\beta(\frac{\lambda_n}{2^n},j\frac{\lambda_n}{2^n})$$

$$=\Gamma(2^{-k}\lambda_n)[\Gamma(\frac{2^{-k}\lambda_n}{2})]^{-2}\int_I r^{\frac{2^{-k}\lambda_n}{2}-1}(1-r)^{\frac{2^{-k}\lambda_n}{2}-1}dr$$

$$=\Gamma(2^{n-k-1}\lambda)[\Gamma(\frac{2^{n-k-1}\lambda}{2})]^{-2}\int_I r^{\frac{2^{n-k-1}}{2}-1}(1-r)^{\frac{2^{n-k-1}}{2}-1}dr$$

$$=\Gamma(\lambda_{n-k})[\Gamma(\frac{\lambda_{n-k}}{2})]^{-2}\int_I r^{\frac{\lambda_{n-k}}{2}-1}(1-r)^{\frac{\lambda_{n-k}}{2}-1}dr$$

$$=Prob^{(n-k)}(E(p_i)\in I)$$

□

3.4 Total Improbability and Marginality

A rather natural question is whether or not there exists a hierarchy of multiplicative priors satisfying both Total Improbability and Marginality. Now it follows trivially from theorem 51 and corollary 50 that there is no hierarchy of priors $\{f^{(n)}\}$ satisfying, Total Improbability and Marginality such that $\forall n>0$ $f^{(n)}\in\mathcal{C}^2_M$ and satisfies Relative Ignorance and Non Nullity. Furthermore, we will show that a similar result is true where the twice differentiability assumption is replaced by the assumption that g is continuous on $[0,1]$.

Lemma 57 *Let* $\{f^{(n)}\}$ *be a hierarchy of multiplicative priors such that* $\forall n>0\ f^{(n)} \in \mathcal{C}_M$ *and satisfies Non Nullity and Weak Renaming. Then given the conditionalisation assumption if* $\{f^{(n)}\}$ *satisfies Marginality and Total Improbability we have that* $\forall n>0\ \forall \vec{x} \in Int(V^{(n)})\ f^{(n)}(\vec{x}) = \frac{\prod_{i=1}^{2^n} g(x_i)}{\int_{V^{(n)}} \prod_{i=1}^{2^n} g(x_i) dV^{(n)}}$ *for some* $g:[0,1] \to \mathbb{R}^{>0}$ *continuous on* $(0,1)$ *such that*

$$\forall x \in (0,1)\ g*g(x) = ke^{cx}g(x)$$

for $k \in \mathbb{R}^{>0}$ *and* $c \in \mathbb{R}$

Proof

Since $\forall n>0\ f^{(n)} \in \mathcal{C}_M$ and $f^{(n)}$ satisfies Non Nullity and $\{f^{(n)}\}$ satisfies Marginality then by lemma 46 $\forall n>0\ f^{(n)}$ satisfies Strong Non Nullity.

Then by Total Improbability and theorem 51

$$\forall \vec{x} \in Int(V^{(n)})\ f^{(n)}(\vec{x}) = \frac{\prod_{i=1}^{2^n} g(x_i)}{\int_{V^{(n)}} \prod_{i=1}^{2^n} g(x_i) dV^{(n)}}$$

where $g:[0,1] \to \mathbb{R}^{\geq 0}$ and is continuous and strictly positive on $(0,1)$ since $\forall n>0\ f^{(n)} \in \mathcal{C}_M$ and satisfies Strong Non Nullity.

Also, by Marginality and lemma 41

$$\forall \vec{x} \in Int(V^{(n)})\ f^{(n)}(\vec{x}) = \frac{\prod_{i=1}^{2^n} g*g(x_i)}{\int_{V^{(n)}} \prod_{i=1}^{2^n} g*g(x_i) dV^{(n)}}$$

Then by continuity, Strong Non Nullity and proposition 27

$$\forall x \in (0,1)\ g * g(x) = ke^{cx} g(x)$$

for $k \in \mathbb{R}^{>0}$ and $c \in \mathbb{R}$ $\Box$

Proposition 58 *Given the conditionalisation assumption there is no hierarchy of priors* $\{f^{(n)}\}$ *satisfying Marginality and Total Improbability such that* $\forall n > 0\ \forall \vec{x} \in V^{(n)}\ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g_i^{(n)}(x_i)$ *where* g_i *is continuous on* $[0,1]$ *for* $i = 1, \ldots, 2^n$ *and* $f^{(n)}$ *satisfies Weak Renaming, Non Nullity and Relative Ignorance.*

Proof

Given continuity, Non Nullity, Weak Renaming and Marginality we have by lemma 46 and lemma 44

$$\forall \vec{x} \in V^{(n)}\ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g^{(n)}(x_i)$$

where $g : [0,1] \to \mathbb{R}^{\geq 0}$ and $f^{(n)}$ satisfies Strong Non Nullity

Now, given Relative Ignorance, by repeating the argument of lemma 49 but substituting $x = 1$ instead of $x = \frac{1}{2}$ we obtain $\forall x, y \in [0,1]$

$$xg^{(n)}(xy)g^{(n)}(x(1-y)) \propto g^{(n)} * g^{(n)}(x)g^{(n)}(y)g^{(n)}(1-y)$$

But by Total Improbability and Marginality, lemma 57 and the continuity of $g^{(n)}$ on $[0,1]$ we have that

$$\forall x \in [0,1]\ g^{(n)} * g^{(n)}(x) = ke^{cx} g^{(n)}(x)$$

where $k \in \mathbb{R}^{>0}$ and $c \in \mathbb{R}$

$\therefore$

$$xg^{(n)}(xy)g^{(n)}(x(1-y)) = k' e^{cx} g^{(n)}(x) g^{(n)}(y) g^{(n)}(1-y)$$

where $k' \in \mathbb{R}^{>0}$

Now, given Strong Non Nullity, we have that $\forall x, y \in (0,1)$

$$g^{(n)}(xy) = \frac{k' e^{cx} g^{(n)}(x)}{xg^{(n)}(x(1-y))} g^{(n)}(y) g^{(n)}(1-y)$$

Hence, interchanging the values of x and y gives

$$\frac{ke^{cx} g^{(n)}(x)}{xg^{(n)}(x(1-y))} g^{(n)}(y) g^{(n)}(1-y) = \frac{ke^{cy} g^{(n)}(y)}{yg^{(n)}(y(1-x))} g^{(n)}(x) g^{(n)}(1-x)$$

$$\Rightarrow\ \forall x, y \in [0,1]\ e^{cx} g^{(n)}(1-y) y g^{(n)}(y(1-x)) = xg^{(n)}(x(1-y)) e^{cy} g^{(n)}(1-x)$$

Letting $x = 0$ we obtain $\forall y \in [0,1]$

$$g^{(n)}(1-y) y g^{(n)}(y) = 0$$

$\therefore$ by continuity $g^{(n)}$ is zero on some interval of $[0,1]$ contradicting Non Nullity

$\square$

A fundamental question remains, however, as to whether there is a hierarchy of multiplicative priors satisfying only Marginality and Total Improbability.

It might be suspected that there exists some function g' on $\mathbb{R}^{\geq 0}$ such that $\int_0^1 g'(x)dx > 0$ and

$$\forall x \in \mathbb{R}^{\geq 0} \; g' * g'(x) = ke^{cx}g'(x)$$

and that a hierarchy of multiplicative priors satisfying Marginality and Total Improbability could be constructed by restricting g' to $[0,1]$. The following proposition shows that this is not the case if g' is integrable on $[0,t]$ for all $t > 0$ and that $\lim_{t\to\infty} \int_0^t g'(x)e^{-ax}dx$ exists for some $a \in \mathbb{R}$

Lemma 59 *If* $L : \mathbb{R}^{\geq a} \to \mathbb{R}$ $a \in \mathbb{R}$ *and* $\forall s \in \mathbb{R}^{>0}$ $L(s) \neq 0$ *then* $\forall s \in \mathbb{R}^{>0}$ $L^2(s) = L(s-c)$ *for* $c \in \mathbb{R}$ *iff* $L(s) = \exp(H(s)2^{-\frac{s}{c}})$ *where* $H : \mathbb{R}^{\geq a} \to \mathbb{R}$ *is a periodic function with period* c*.*

Proof

There are three cases to consider

(i) Suppose $L(s)$ is defined $\forall s > a$, $c > a$ and $L(c) \neq 1$

$\therefore \; L(s) = L^2(s+c) \;\Rightarrow\; L(s) > 0$

Then taking logs and letting $f = \log L$ gives $2f(s+c) = f(s)$

Now since $L(c) \neq 1$ then $f(c) \neq 0$ which implies $\frac{f(s)}{f(c)}$ exist for all $s > a$

Now make the transformation $R(s) = \frac{f(s)}{2f(c)}$ to obtain $R(s+c) = R(s)R(c)$.

From this we obtain the following recursive formula

$$R((n-1)c+c) = R((n-1)c)R(c) \;\Rightarrow\; R(nc) = (R(c))^n$$

Now $R(c) = \frac{f(c)}{2f(c)} = \frac{1}{2} > 0$ and $\therefore \log(R(c))$ exists. Taking logs then we obtain

$$\log(R(nc)) = n\log(R(c)) = nc\frac{\log(R(c))}{c}$$

Thus letting $k = \frac{\log(R(c))}{c}$ then $R(nc) = e^{(nc)k}$

Now let $\tau(s) = R(s)e^{-sk}$ then clearly τ is periodic with period c and

$$R(s) = \tau(s)e^{sk} = \tau(s)e^{\frac{s}{c}\log(R(c))} = \tau(s)e^{-\frac{s}{c}\log 2} = \tau(s)2^{-\frac{s}{c}}$$

Then letting $H(s) = 2f(c)\tau(s)$ we obtain $f(s) = H(s)2^{-\frac{s}{c}}$ which implies $L(s) = \exp(H(s)2^{-\frac{s}{c}})$

(ii) Suppose $L(c) = L(nc) = 1$ then let $M(s) = e^{2^{-\frac{s}{c}}}L(s)$. Clearly $M(s) = M^2(s+c)$ and $M(c) = e^2 \neq 1$

Then by the above argument $M(s) = \exp(H^*(s)2^{-\frac{s}{c}})$ for a periodic function H^* with period c

$$\therefore L(s) = M(s).e^{-2^{-\frac{s}{c}}} = \exp((H^*(s)-1)2^{-\frac{s}{c}}) = \exp(H(s)2^{-\frac{s}{c}})$$

where $H(s) = H^*(s) - 1$

(iii) If $L(s)$ is defined only for $s > a$ and $c \leq a$ then let $s = t + a \quad t > 0$ then $L(t+a) = L^2(t+a+c)$. Substituting $M(t) = L(t+a)$ gives $M(t) = M^2(t+c)$ then as above $M(t) = \exp(H^*(t)2^{-\frac{t}{c}})$ where H^* is a periodic function with period c

$$L(s) = \exp(H^*(s-a)2^{-\frac{s}{c}}2^{\frac{a}{c}}) = \exp(H(s)2^{-\frac{s}{c}})$$

where $H(s) = H^*(s-a)2^{\frac{a}{c}}$ $\Box$

Proposition 60 *There is no function $g' : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ such that g' is integrable on $[0,t]$ for all $t > 0$, $\int_0^1 g'(t)dt > 0$, $\lim_{t\to\infty} \int_0^t g'(x)e^{-ax}dx$ exists for some real constant a and $\forall x > 0$ $g' * g'(x) = ke^{cx}g'(x)$*

Proof

If g' has the above properties then it has a Laplace transform $L(s)$ such that $\forall s > a$ $L^2(s) = L(s-c)$ and $L(s) \neq 0$.

Then by the previous lemma it suffices to show that $L(s) = \exp(H(s)2^{-\frac{s}{c}})$ is not a Laplace transform of a function.

If $L(s)$ is a Laplace transform then $L(s) \to 0$ as $s \to \infty$ (see [24]) $\therefore$ since $H(s)$ is periodic and defined for all $s > 0$ it follows that $H(s)2^{-\frac{s}{c}} \to 0$ as $s \to \infty$

$\therefore$ $L(s) \to 1$ as $s \to \infty$ $\Box$

Theorem 61 *Given the conditionalisation assumption there is no hierarchy of priors $\{f^{(n)}\}$ satisfying Marginality and Total Improbability such that $\forall n > 0$ $f^{(n)} \in \mathcal{C}_M^\infty$ and satisfies Weak Renaming and Non Nullity*

Proof

Given Weak Renaming, Non Nullity, Marginality and Total Improbability we have by lemma 57 that $\forall n > 0$ $\forall \vec{x} \in Int(V^{(n)})$ $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$ where g

satisfies

$$\forall x \in (0,1)\ g * g(x) = ke^{cx}g(x)$$

Further, since $f^{(n)} \in \mathcal{C}_M^\infty$ $\exists a \in (0,1)$ such that $g(x) = \sum_{n=0}^\infty \frac{g^{[n]}(a)}{n!}(x-a)^n$ where $g^{[n]}(a) = \frac{d^n}{dx^n}g(x)|_{x=a}$

Therefore

$$g * g(x) = \int_0^x g(t)g(x-t)dt = \int_0^x \sum_{n=0}^\infty \sum_{m=0}^\infty \frac{g^{[n]}(a)}{n!}\frac{g^{[m]}(a)}{m!}(t-a)^n(x-t-a)^m dt$$

Now $(t-a)^n = \sum_{r=0}^n C_r^n a^{n-r}t^r(-1)^{n-r}$

and similarly $(x-t-a)^m = \sum_{s=0}^m C_s^m a^{m-s}(-1)^{m-s}(x-t)^s$ $\therefore$

$$g*g(x) = \int_0^x \sum_{n=0}^\infty \sum_{m=0}^\infty \frac{g^{[n]}(a)}{n!}\frac{g^{[m]}(a)}{m!}\sum_{r=0}^n \sum_{s=0}^m C_r^n C_s^m a^{n+m-r-s}(-1)^{n+m-r-s}t^r(x-t)^s dt$$

$$= \sum_{n=0}^\infty \sum_{m=0}^\infty g^{[n]}(a)g^{[m]}(a)\sum_{r=0}^n \sum_{s=0}^m \frac{a^{n+m-r-s}(-1)^{n+m-r-s}}{r!(n-r)!s!(m-s)!}\int_0^x t^r(x-t)^s dt$$

$$= \sum_{n=0}^\infty \sum_{m=0}^\infty g^{[n]}(a)g^{[m]}(a)\sum_{r=0}^n \sum_{s=0}^m \frac{a^{n+m-r-s}(-1)^{n+m-r-s}}{r!(n-r)!s!(m-s)!}x^{r+s+1}\beta(r+1,s+1)$$

$$= \sum_{n=0}^\infty \sum_{m=0}^\infty g^{[n]}(a)g^{[m]}(a)\sum_{r=0}^n \sum_{s=0}^m \frac{(-1)^{n+m-r-s}a^{n+m-r-s}x^{r+s+1}}{(n-r)!(m-s)!(r+s+1)!}$$

Also

$$e^{cx}g(x) = e^{cx}\sum_{n=0}^\infty \frac{g^{[n]}(a)}{n!}\sum_{r=0}^n x^r a^{n-r}(-1)^{n-r}C_r^n$$

$$= \sum_{m=0}^\infty \frac{c^m x^m}{m!}\sum_{n=0}^\infty \frac{g^{[n]}(a)}{n!}\sum_{r=0}^n x^r a^{n-r}(-1)^{n-r}C_r^n = \sum_{m=0}^\infty \sum_{n=0}^\infty \frac{c^m}{m!}g^{[n]}(a)\sum_{r=0}^n \frac{a^{n-r}(-1)^{n-r}}{r!(n-r)!}x^{r+m}$$

Therefore, we have that

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} g^{[n]}(a)g^{[m]}(a)\sum_{r=0}^{n}\sum_{s=0}^{m}\frac{(-1)^{n+m-r-s}a^{n+m-r-s}x^{r+s+1}}{(n-r)!(m-s)!(r+s+1)!}$$

$$=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{c^m}{m!}g^{[n]}(a)\sum_{r=0}^{n}\frac{a^{n-r}(-1)^{n-r}}{r!(n-r)!}x^{r+m}$$

We now show by induction on k that $\sum_{n=k}^{\infty} g^{[n]}(a)\frac{a^{n-k}(-1)^{n-k}}{(n-k)!} = 0$

Comparing coefficients of x^k then for $k=0$ we obtain $\sum_{n=0}^{\infty} g^{[n]}(a)\frac{a^n(-1)^n}{n!} = 0$ and for $k \geq 1$ we have

$$\sum_{m=0}^{k}\sum_{n=k-m}^{\infty}\frac{c^m}{m!}g^{[n]}(a)\frac{a^{n+m-k}(-1)^{n+m-k}}{(k-m)!(n+m-k)!}$$
$$=\sum_{s=0}^{k-1}\sum_{m=s}^{\infty}\sum_{n=k-s-1}^{\infty}\frac{(-1)^{n+m-k+1}a^{n+m-k+1}}{(n-k+s+1)!(m-s)!k!}g^{[n]}(a)g^{[m]}(a)$$

Now,

$$R.H.S = \frac{1}{k!}\sum_{s=0}^{k-1}\sum_{m=s}^{\infty}\frac{(-1)^{m-s}a^{m-s}g^{[m]}(a)}{(m-s)!}\sum_{n=k-s-1}^{\infty}\frac{(-1)^{n-(k-s-1)}a^{n-(k-s-1)}}{(n-k+s+1)!}g^{[n]}(a) = 0$$

by the inductive step

$$L.H.S = \sum_{n=k}^{\infty}\frac{g^{[n]}(a)a^{n-k}(-1)^{n-k}}{k!(n-k)!} + \sum_{m=1}^{k}\frac{c^m}{m!}\sum_{n=k-m}^{\infty}g^{[n]}(a)\frac{a^{n-k+m}(-1)^{n-k+m}}{(k-m)!(n+m-k)!}$$
$$=\sum_{n=k}^{\infty}g^{[n]}(a)\frac{a^{n-k}(-1)^{n-k}}{k!(n-k)!}$$

$\therefore$ we have $\forall x \in (0,1)$ $g * g(x) = 0$ since all coefficients of x^k for $k \geq 0$ are zero in the expansion of $g * g$ $\therefore \forall x \in (0,1)$ $g(x) = 0$ $\Box$

Chapter 4

Centre of Mass

In this chapter Centre of Mass is considered in the context of inference processes and a number of the axioms on priors proposed above are related to the principles on inference processes discussed in Chapter 1. In addition, it will be shown that there exist prior densities $f^{(n)} \in \mathcal{C}$ satisfying Strong Non Nullity for which $CM^{f^{(n)}}$ on equality knowledge bases does not form an equality inference process. That is, for these priors there is some $K \in EKB$ such that $CM_K^{f^{(n)}}$ is not defined. If, however, we restrict ourselves to open interval knowledge bases this problem does not arise.

4.1 Centre of Mass on Open Interval Knowledge Bases

Proposition 62 *If $f^{(n)}$ is a prior density on $V^{(n)}$ satisfying Non Nullity then $CM^{f^{(n)}}$ on open interval knowledge bases is an open interval inference process*

Proof

By Non Nullity and corollary 18 $\forall K \in IKB(L^{(n)}) \int_{V^{(n)}(K)} f^{(n)}(\vec{x})dV^{(n)} > 0$

Hence, $\forall \vec{x} \in V^{(n)}(K)$ $f^{(n)}(\vec{x}|K)$ is defined. $\Box$

Proposition 63 *If $h(\vec{x}), f(\vec{x}) \in \mathcal{C}$ are density functions on $V^{(n)}$ satisfying Strong Non Nullity such that*

$$\forall \theta \in SL^{(n)} \ CM_K^f(\theta) = CM_K^h(\theta)$$

for all open interval knowledge bases of $L^{(n)}$ then

$$\forall \vec{x} \in IntV^{(n)} \quad h(\vec{x}) = \kappa f(\vec{x})$$

for $\kappa \in \mathbb{R}^{>0}$

Proof

For $\vec{t} \in Int(V^{(n)})$ let $I_1, \ldots, I_{2^n-1}$ be open intervals of $V^{(n)}$ such that $I_i = (t_i - \delta, t_i + \delta)$ for $i = 2, \ldots, 2^n - 1$. Then for $0 < z < y < 1 - \sum_{i=2}^{2^n-1} t_i$ if $I_1 = (z, y)$ then $\forall \delta > 0$ $V^{(n)}(\{E(\alpha_i^{(n)}) \in I_i \ , i = 1, \ldots 2^n - 1\}) \neq \emptyset$. Hence, we obtain

$$\lim_{\delta \to 0} \frac{\int_{I_1} \int_{I_2} \cdots \int_{I_{2^n-1}} x_1 f(\vec{x}) dx_1 \ldots dx_{2^n-1}}{\int_{I_1} \int_{I_2} \cdots \int_{I_{2^n-1}} f(\vec{x}) dx_1 \ldots dx_{2^n-1}} = \lim_{\delta \to 0} \frac{\int_{I_1} \int_{I_2} \cdots \int_{I_{2^n-1}} x_1 h(\vec{x}) dx_1 \ldots dx_{2^n-1}}{\int_{I_1} \int_{I_2} \cdots \int_{I_{2^n-1}} h(\vec{x}) dx_1 \ldots dx_{2^n-1}}$$

Now by Strong Non Nullity and the continuity of $f^{(n)}$ on $Int(V^{(n)})$ we have , taking the limit as δ tends to zero, that

$$\frac{\int_{I_1} x_1 f(x_1, t_2, \ldots, t_{2^n-1}, 1 - x_1 - \sum_{i=2}^{2^n-1} t_i) dx_1}{\int_{I_1} f(x_1, t_2, \ldots, t_{2^n-1}, 1 - x_1 - \sum_{i=2}^{2^n-1} t_i) dx_1} =$$

$$\frac{\int_{I_1} x_1 h(x_1, t_2, \ldots, t_{2^n-1}, 1 - x_1 - \sum_{i=2}^{2^n-1} t_i) dx_1}{\int_{I_1} h(x_1, t_2, \ldots, t_{2^n-1}, 1 - x_1 - \sum_{i=2}^{2^n-1} t_i) dx_1}$$

Letting $H_{\vec{t}}(x) = h(x, t_2, \ldots, t_{2^n-1}, 1 - x - \sum_{i=2}^{2^n-1} t_i)$

and $F_{\vec{t}}(x) = f(x, t_2, \ldots, t_{2^n-1}, 1 - x - \sum_{i=2}^{2^n-1} t_i)$ and $I_1 = (z, y)$

Then differentiating with respect to y gives

$$yF_{\vec{t}}(y)\int_{(z,y)} H_{\vec{t}}(x)dx + H_{\vec{t}}(y)\int_{(z,y)} xF_{\vec{t}}(x)dx = F_{\vec{t}}(y)\int_{(z,y)} xH_{\vec{t}}(x)dx + yH_{\vec{t}}(y)\int_{(z,y)} F_{\vec{t}}(x)dx$$

Now differentiating with respect to z gives

$$yF_{\vec{t}}(y)H_{\vec{t}}(z) \ + \ H_{\vec{t}}(y)zF_{\vec{t}}(z)$$

$$= \ F_{\vec{t}}(y)zH_{\vec{t}}(z) \ + \ yH_{\vec{t}}(y)F_{\vec{t}}(z)$$

$$\Rightarrow \ F_{\vec{t}}(y)H_{\vec{t}}(z)(y-z) = H_{\vec{t}}(y)F_{\vec{t}}(z)(y-z)$$

$$\Rightarrow \ \frac{F_{\vec{t}}(y)}{H_{\vec{t}}(y)} = \frac{F_{\vec{t}}(z)}{H_{\vec{t}}(z)}$$

Therefore, since z can be taken to be arbitrarily close to 0 we have

$$\forall y \in (0, 1 - \sum_{i=2}^{2^n-1} t_i) \ \ F_{\vec{t}}(y) = \kappa'(t_2, \ldots t_{2^n-1})H_{\vec{t}}(y)$$

for $\kappa' : \mathbb{R}^{2^n-2} \to \mathbb{R}^{>0}$

$$\Rightarrow \ f(\vec{t}) = \kappa'(t_2, \ldots t_{2^n-1})h(\vec{t})$$

Now repeating the above argument varying the 2'nd to the $2n-1$'th coordinates in turn gives

$$\forall \vec{t} \in Int(V^{(n)})f(\vec{t}) = \kappa h(\vec{t})$$

for some $\kappa \in \mathbb{R}^{>0}$ □

Corollary 64 *If $\{f^{(n)}\}$ is a hierarchy of priors such that $\forall n > 0$ $f^{(n)} \in \mathcal{C}$ and $f^{(n)}$ satisfies Strong Non Nullity then*

$\{f^{(n)}\}$ satisfies Marginality iff $\{CM^{f^{(n)}}\}$ satisfies Language Invariance

Proof

($\Rightarrow$)

Follows trivially

($\Leftarrow$)

For all open interval knowledge bases K of $L^{(n-1)}$

$$CM_K^{f^{(n-1)}}(\alpha_j^{(n-1)}) = \frac{\int_{a_1}^{b_1} \dots \int_{a_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})}^{b_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})} y_j f^{(n-1)}(\vec{y})d\vec{y}}{\int_{a_1}^{b_1} \dots \int_{a_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})}^{b_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})} f^{(n-1)}(\vec{y})d\vec{y}}$$

where the a_j's are the maximum of some set of linear functions and the b_j's are the minimum of some set of linear functions

Then after making a change of variables to x_{2i-1}, y_i where $y_i = x_{2i-1} + x_{2i}$ for $i = 1, \dots, 2^{n-1}$ we have

$$CM_K^{f^{(n)}}(\alpha_j^{(n-1)}) =$$

$$\frac{\int_{a_1}^{b_1} \dots \int_{a_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})}^{b_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})} y_j h^{(n-1)}(\vec{y})d\vec{y}}{\int_{a_1}^{b_1} \dots \int_{a_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})}^{b_{2^n-1}(y_1,\dots,y_{2^{n-1}-2})} h^{(n-1)}(\vec{y})d\vec{y}}$$

where

$$h^{(n-1)}(\vec{y}) = \int_0^{y_1} \dots \int_0^{y_{2^n-1}} f^{(n)}(x_1, y_1 - x_1, \dots, x_{2^n-1}, y_{2^{n-1}} - x_{2^n-1})d\vec{x}$$

$$= CM_K^{h^{(n-1)}}(\alpha_j^{(n-1)})$$

$\Rightarrow$ $\{f^{(n)}\}$ satisfies Marginality by proposition 63 $\Box$

Corollary 65 *If $f^{(n)} \in \mathcal{C}$ satisfying Strong Non Nullity then $f^{(n)}$ satisfies Renaming iff $CM^{f^{(n)}}$ on open interval knowledge bases satisfies Invariance under Renaming*

Proof

Since any automorphism on $\overline{SL^{(n)}}$ is uniquely determined by it's values on the atoms of $\overline{SL^{(n)}}$ it follows that any such automorphism σ uniquely defines a permutation of $\{1, \ldots, 2^n\}$ given by $\sigma(i) = j$ iff $\sigma(\bar{\alpha}_i^{(n)}) = \bar{\alpha}_j^{(n)}$ and vica versa.

Now $CM^{f^{(n)}}$ on open interval knowledge bases satisfies Invariance under Renaming iff $\forall K \in IKB$

$$\frac{\int_{V^{(n)}(K)} x_j f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}(K)} f^{(n)}(\vec{x}) dV^{(n)}} = \frac{\int_{V^{(n)}(K_\sigma)} x_{\sigma(j)} f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}(K_\sigma)} f^{(n)}(\vec{x}) dV^{(n)}}$$

for all automorphisms $\sigma : \overline{SL^{(n)}} \to \overline{SL^{(n)}}$

Then making the change of variables to $y_j = x_{\sigma(j)}$ the R.H.S equals

$$\frac{\int_{V^{(n)}(K)} y_j f^{(n)}(\sigma(\vec{y})) dV^{(n)}}{\int_{V^{(n)}(K)} f^{(n)}(\sigma(\vec{y})) dV^{(n)}}$$

($\Rightarrow$)

Follows immediately

$(\Leftarrow)$

It follow from proposition 63 that

$$\forall \vec{x} \in Int(V^{(n)})\ f^{(n)}(\vec{x}) = k_\sigma f^{(n)}(\sigma(\vec{x}))$$

where $k_\sigma \in \mathbb{R}^{>0}$

Now $\int_{V^{(n)}} f^{(n)}(\vec{x}) dV^{(n)} = 1$ and making the change variable to $y_i = \sigma(x_i)$ we have $\int_{V^{(n)}} f^{(n)}(\sigma(\vec{x})) dV^{(n)} = 1 \Rightarrow\ k_\sigma = 1$ $\Box$

4.2 Centre of Mass on Equality Knowledge Bases

In this section we consider Centre of Mass relative to Dirichlet priors on equality knowledge bases. We will see that there are problems with the definition given in that for some relatively simple knowledge bases the center of mass is undefined. Initially, however, we again consider the principles of Language Invariance and ZP Invariance. Naturally throughout this section we will be assuming the conditionalisation assumption.

Definition 66 *For* $K \in EKB(L^{(n)})$ $Inc_n(K) = \{i \mid \forall \vec{x} \in V^{(n)}(K), x_i = 0\}$

Definition 67 *If* $f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i) \in \mathcal{C}_M$ *then for* $K \in EKB(L^{(n)})$ *and* $V(K) \subseteq B(V^{(n)})$

$$CM_K^{f^{(n)}}(\alpha_i^{(n)}) = \frac{\int_{V^{(n)}(K)} x_i \prod_{i \notin Inc_n(K)} g(x_i) dV^{(n)}(K)}{\int_{V^{(n)}(K)} \prod_{i \notin Inc_n(K)} g(x_i) dV^{(n)}(K)}$$

if $\prod_{i\notin Inc_n(K)} g_i(x_i)$ *is integrable on* $V^{(n)}(K)$ *and* $\int_{V^{(n)}(K)} \prod_{i\notin Inc_n(K)} g(x_i)dV^{(n)}(K) > 0$ *and is left undefined otherwise*

Note that this definition corresponds to the conditionalisation assumption in the case where g is strictly positive on $[0, 1]$.

Notation : For $f^{(n)} \in \mathcal{C}$ and $K \in EKB(L^{(n)})$ $f^{(n)} \restriction V^{(n)}(K)$ will be used to denote the function obtained by restricting $f^{(n)}$ to $V^{(n)}(K)$.

Proposition 68 *If* $f^{(n)} \in \mathcal{C}_M$ *is a symmetric multiplicative prior continuous on* $V^{(n)}$ *satisfying Strong Non Nullity then* $CM^{f^{(n)}}$ *on equality knowledge bases is an equality inference process.*

Proof

Let K be a consistent equality knowledge base on $L^{(n)}$ such that $V^{(n)}(K) \not\subseteq B(V^{(n)})$. Now by Strong Non Nullity $\forall \vec{x} \in Int(V^{(n)}K)$ $f^{(n)}(\vec{x}) > 0$. Also $f^{(n)}$ is continuous on $V^{(n)}$ which implies that $f^{(n)} \restriction V^{(n)}(K)$ is integrable on $V^{(n)}(K)$ and $\int_{V^{(n)}(K)} f^{(n)}(\vec{x})dV^{(n)}(K) > 0$

Alternatively let $K \in EKB(L^{(n)})$ be such that $V^{(n)}(K) \subseteq B(V^{(n)})$. Now since $f^{(n)}$ is a symmetric multiplicative prior continuous on $V^{(n)}$ we have

$$\forall \vec{x} \in V^{(n)} \ f^{(n)}(\vec{x}) = \prod_{i=1}^{2^n} g(x_i)$$

for some $g : [0, 1] \to \mathbb{R}^{\geq 0}$ where g is continuous on $[0, 1]$. Also by Strong Non Nullity $\forall x \in (0, 1)$ $g(x) > 0$. Hence, $\int_{V^{(n)}(K)} \prod_{i\notin Inc_n(K)} g(x_i)dV^{(n)}(K)$ exists and is non zero and hence by definition 67 $CM_K^{f^{(n)}}$ exists $\square$

An obvious consequence of definition 67 is that for hierarchies of priors where for all $n > 0$ $f^{(n)}$ satisfies the conditions of proposition 68 and $\{f^{(n)}\}$ satisfies Total Improbability then $\{CM^{f^{(n)}}\}$ on equality knowledge bases is a hierarchy of inference processes satisfying ZP Invariance.

Now it can easily be seen that there is no hierarchy of Dirichlet priors $\{d(\lambda_n, n)\}$ for which $\{CM^{d(\lambda_n,n)}\}$ restricted to equality knowledge bases is a hierarchy of inference processes satisfying ZP Invariance and Language Invariance. Hence, there is no such hierarchy satisfying Irrelevant Information.

Notice that any inference process on equality knowledge bases which selects that $\vec{x} \in V^{(n)}(K)$ maximising $\sum_{i \notin Inc_n(K)} f(y_i)$ for some convex function f satisfies ZP Invariance. An example of such an inference process can be derived from $CM^{d(\lambda,n)}$ by letting λ tend to infinity.

Proposition 69 *For all knowledge bases* $K \in EKB(L^{(n)})$ *and* $\forall \theta \in SL^{(n)}$

$CM_K^{\infty}(L^{(n)})(\theta) = \lim_{\lambda \to \infty} CM_K^{d(\lambda,n)}(\theta)$ *exists and* $CM_K^{\infty}(\alpha_i^{(n)}) = x_i$ *where*

$\vec{x} = Max_{\vec{y} \in V^{(n)}(K)} \sum_{i \notin Inc_n(K)} \ln y_i$

Proof

For $\lambda > 2^n$ $d(\lambda, n)$ is continuous on $V^{(n)}$ and hence, $\prod_{i \notin Inc_n(K)} x_i^{\frac{\lambda}{2^n}-1}$ is integrable on $V^{(n)}(K)$.

Also, for $\lambda > 2^n$ $\prod_{i \notin Inc_n(K)} x_i^{\frac{\lambda}{2^n}-1} \restriction V^{(n)}(K)$ is a convex function which becomes increasingly peaked at its maximum as λ tends to infinity. Hence,

$$\lim_{\lambda\to\infty} CM_K^{d(\lambda,n)}(\alpha_i^{(n)}) = \lim_{\lambda\to\infty} \frac{\int_{V^{(n)}(K)} x_i \prod_{i\notin Inc_n(K)} x_i^{\frac{\lambda}{2^n}-1} dV^{(n)}(K)}{\int_{V^{(n)}(K)} \prod_{i\notin Inc_n(K)} x_i^{\frac{\lambda}{2^n}-1} dV^{(n)}(K)} = x_i$$

where $\vec{x} = Max_{\vec{y}\in V^{(n)}(K)} \sum_{i\notin Inc_n(K)} \ln y_i$ $\square$

The following is a generalisation of a result due to Paris and Vencovska (see [18]) where $\{f^{(n)}\}$ was assumed to be the hierarchy of uniform priors.

Proposition 70 *If* $\{f^{(n)}\} = \{d(\kappa 2^n, n)\}$ *for some* $\kappa \in \mathbb{R}^{\geq 0}$ *then for all knowledge bases* $K \in EKB(L^{(n)})$ *and* $\forall\theta \in SL^{(n)}$

$\lim_{m\to\infty} CM_K^{f^{(n+m)}}(\theta)$ *exists and equals* $CM_K^{\infty}(L^{(n)})(\theta)$

Proof

Let K be a equality knowledge base where $dim(V^{(n)}(K)) = s$. w.l.o.g assume that $V^{(n)}(K)$ is parameterised by $x_1, \ldots, x_s$ so that these are free variables and $x_i = l_i(x_1, \ldots, x_s)$ for $i > s$ where l_i are linear functions. Then for $f^{(n)}$ a symmetric multiplicative prior on $V^{(n)}$ integrable on $V^{(n)}(K)$

$$CM_K^{f^{(n)}}(\alpha_i^{(n)}) = \frac{\int_{a_1}^{b_1}\int_{a_2(x_1)}^{b_2(x_1)} \cdots \int_{a_s(x_1\ldots x_{s-1})}^{b_s(x_1\ldots x_{s-1})} x_i \prod_{i\notin Inc_n(K)} g^{(n)}(x_i) dx_1 \ldots dx_s}{\int_{a_1}^{b_1}\int_{a_2(x_1)}^{b_2(x_1)} \cdots \int_{a_s(x_1\ldots x_{s-1})}^{b_s(x_1\ldots x_{s-1})} \prod_{i\notin Inc_n(K)} g^{(n)}(x_i) dx_1 \ldots dx_s}$$

where the a_j's are the maximum of some set of linear functions and the b_j's are the minimum of some set of linear functions

We now extend the language by adding the m propositional variables of $L^{(m+n)} - L^{(n)}$. Let $\beta_j^{(m)}$ for $j = 1, \ldots, 2^m$ be an enumeration of the atoms of this language so that $\alpha_{i,j}^{(n+m)} \equiv \alpha_i^{(n)} \wedge \beta_j^{(m)}$ is an enumeration of the atoms of $L^{(n+m)}$

Let $x_{i,j} = E(\alpha_{i,j}^{n+m})$ and make the change of variables $x_i = \sum_{j=1}^{2^m} x_{i,j}$, $y_{i,1} = \sum_{j=1}^{2^m} x_{i,j}$, . . . , $y_{i,2^m-1} = x_{i,1}$

Then $\forall \vec{x} \in Int(V^{(n+m)})$

$$d(\lambda_{n+m}, n+m)(\vec{x}) \propto \prod_{i=1}^{2^n} \prod_{j=1}^{2^m} x_{i,j}^{\frac{\lambda_{n+m}}{2^{n+m}}-1}$$

$$\prod_{i=1}^{2^n} (x_i - y_{i,1})^{\frac{\lambda_{n+m}}{2^{n+m}}-1} y_{i,2^m-1}^{\frac{\lambda_{n+m}}{2^{n+m}}-1} \prod_{j=2}^{2^m-1} (y_{i,2^m-j} - y_{i,2^m-j+1})^{\frac{\lambda_{n+m}}{2^{n+m}}-1}$$

Thus we see that $CM_K^{d(\lambda_{n+m},n+m)}(\alpha_i^{(n)})$ exists if

$$\prod_{i \notin Inc_n(K)} \int_0^{x_1} \int_0^{y_{i,1}} \cdots \int_0^{y_{i,2^m-2}} (x_1 - y_{i,1})^{\frac{\lambda_{n+m}}{2^{n+m}}-1} y_{i,2^m-1}^{\frac{\lambda_{n+m}}{2^{n+m}}-1}$$
$$\prod_{j=2}^{2^m-1} (y_{i,2^m-j} - y_{i,2^m-j+1})^{\frac{\lambda_{n+m}}{2^{n+m}}-1} dy_{i,1} \ldots dy_{i,2^m-1}$$

$$= [\beta(\lambda_{n+m}, r\lambda_{n+m})]^{2^n} \prod_{i=1}^{2^n} x_i^{\frac{\lambda_{n+m}}{2^n}-1}$$

$$[\beta(\lambda_{n+m}, r\lambda_{n+m})]^{2^n} \prod_{i=1}^{2^n} x_i^{\kappa 2^m-1}$$

is integrable on $V^{(n)}(K)$. Now for $m > \log_2(\frac{1}{\kappa})$ $\prod_{i=1}^{2^n} x_i^{\kappa 2^m-1}$ is continuous on $V^{(n)}$ and hence integrable on $V^{(n)}(K)$ and

$$\therefore\ CM_K^{d(\lambda_{n+m},n+m)}(\alpha_i^{(n)}) = \frac{\int_{a_1}^{b_1}\int_{a_2(x_1)}^{b_2(x_1)}\cdots\int_{a_s(x_1\ldots x_{s-1})}^{b_s(x_1\ldots x_{s-1})} x_i \prod_{i\notin Inc_n(K)} x_i^{\kappa 2^m - 1}}{\int_{a_1}^{b_1}\int_{a_2(x_1)}^{b_2(x_1)}\cdots\int_{a_s(x_1\ldots x_{s-1})}^{b_s(x_1\ldots x_{s-1})} \prod_{i\notin Inc_n(K)} x_i^{\kappa 2^m - 1}}$$

Further for $m > \log_2(\frac{1}{\kappa})$ $\prod_{i=1}^{2^n} x_i^{\kappa 2^m - 1} \restriction V^{(n)}(K)$ is a convex function which becomes increasingly peaked at it's maximum as m tends to infinity giving the required result . $\square$

Proposition 71 $\{CM^\infty(L^{(n)})\}$ *is Language Invariant*

Proof

See Paris [22] $\square$

However, CM^∞ does not satisfy the stronger principle of Irrelevant Information as is illustrated by the following example.

Example

Let $L^{(1)} = \{p_1\}$, $L^{(3)} = \{p_1, p_2, p_3\}$ and $K_1 = \{E(p_1) = \frac{1}{3}\}$, $K_2 = \{E(p_2 \wedge p_3) + \frac{1}{4}E(\neg p_2 \wedge \neg p_3) = \frac{1}{4}, E(p_2 \wedge \neg p_3) = 0\}$ so that $K_1 \in KB(L^{(1)})$ and $K_2 \in KB(L^{(3)} - L^{(1)})$

If β_j for $j = 1, \ldots, 4$ is an enumeration of the atoms of $L^{(3)} - L^{(1)}$ then let $z_{1,j} = E(p_1 \wedge \beta_j)$ and $z_{2,j} = E(\neg p_1 \wedge \beta_j)$. Also, let

$$\vec{z} =< z_{1,1}, z_{2,1}, z_{1,2}, z_{2,2}, z_{1,3}, z_{2,3}, z_{1,4}, z_{2,4} >$$

Then, we have

$K_1 \cup K_2 = \{z_{1,1} + z_{1,2} + z_{1,3} + z_{1,4} = \frac{1}{3}$, $z_{2,1} + z_{2,2} + z_{2,3} + z_{2,4} = \frac{2}{3}$, $z_{1,1} + z_{2,2} + \frac{1}{4}z_{1,4} + \frac{1}{4}z_{2,4} = \frac{1}{4}$, $z_{1,2} + z_{2,2} = 0\}$

$\therefore\ K_1 \cup K_2 \ \Rightarrow z_{1,2} = z_{2,2} = 0$, $z_{1,4} = \frac{1}{3} - z_{1,1} - z_{1,3}$ and $z_{2,4} = \frac{2}{3} - z_{2,1} - z_{2,3}$

also $\frac{1}{4}z_{2,4} = \frac{1}{4} - z_{1,1} - z_{2,1} - z_{2,1} - (\frac{1}{3} - z_{1,1} - z_{1,3})$ and $z_{2,4} = \frac{2}{3} - 3z_{1,1} - 4z_{2,1} + z_{1,3}$

$\therefore\ \frac{2}{3} - z_{2,1} - z_{2,3} = \frac{2}{3} - 3z_{1,1} - 4z_{2,1} + z_{1,3} \ \Rightarrow \ z_{2,3} = 3(z_{1,1} + z_{2,1}) - z_{1,3}$

Hence, a parameterisation of $V^{(3)}(K_1 + K_2)$ is

$$< z_{1,1},\ z_{2,1},\ 0,\ 0,\ z_{1,3},\ \frac{1}{3} - z_{1,1} - z_{1,3},\ \frac{2}{3} - 3z_{1,1} - 4z_{2,1} + z_{2,3},\ 3(z_{1,1} + z_{2,1}) - z_{1,3} >$$

If $L = \sum_{i=1}^{2} \sum_{j \notin Inc_n(K_2)} \log z_{i,j}$ then

$$L(\vec{z}) = \log z_{1,1} + \log z_{1,3} + \log z_{2,1} + \log(\frac{1}{3} - z_{1,1} - z_{1,3}) +$$
$$\log(\frac{2}{3} - 3z_{1,1} - 4z_{2,1} + z_{2,3}) + \log(3(z_{1,1} + z_{2,1}) - z_{1,3})$$

Hence, if $\vec{z} = Max_{\vec{z} \in V^{(3)}(K_1 + K_2)} L(\vec{z})$, $\vec{z}$ is a solution the following simultaneous equations

$$\frac{\partial L}{\partial z_{1,1}} = \frac{1}{z_{1,1}} - \frac{1}{\frac{1}{3} - z_{1,1} - z_{1,3}} - \frac{3}{\frac{2}{3} - 3z_{1,1} + 4z_{2,1} + z_{1,3}} + \frac{3}{3(z_{1,1} + z_{2,1}) - z_{1,3}} = 0$$

$$\frac{\partial L}{\partial z_{1,3}} = \frac{1}{z_{1,3}} - \frac{1}{\frac{1}{3} - z_{1,1} - z_{1,3}} - \frac{3}{\frac{2}{3} - 3z_{1,1} - 4z_{2,1} + z_{1,3}} - \frac{1}{3(z_{1,1} + z_{2,1}) - z_{1,3}} = 0$$

$$\frac{\partial L}{\partial z_{2,1}} = \frac{1}{z_{2,1}} - \frac{4}{\frac{2}{3} - 3z_{1,1} - 4z_{2,1} + z_{1,3}} = 0$$

Now $CM^{\infty}_{K_2}(L^{(3)} - L^{(1)})(p_2 \wedge p_3) = \frac{1}{6}$ hence, for irrelevant information to hold we require that $z_{1,1} + z_{2,1} = \frac{1}{6}$. So substituting this into the above equations we obtain

$$\frac{1}{z_{1,1}} - \frac{1}{\frac{1}{3} - z_{1,1} - z_{1,3}} - \frac{3}{z_{1,1} + z_{1,3}} + \frac{3}{\frac{1}{2} - z_{1,3}} = 0 \quad [1]$$

$$\frac{1}{z_{1,3}} - \frac{1}{\frac{1}{3} - z_{1,1} - z_{1,3}} + \frac{1}{z_{1,1} + z_{1,3}} - \frac{1}{\frac{1}{2} - z_{1,3}} = 0 \quad [2]$$

$$\frac{1}{\frac{1}{6} - z_{1,1}} - \frac{4}{z_{1,1} + z_{1,3}} + \frac{3}{\frac{1}{2} - z_{1,3}} = 0 \quad [3]$$

Solving these equation simultaneously $[1] + 3[2]$ gives

$$\frac{1}{z_{1,1}} + \frac{3}{z_{1,3}} - \frac{4}{\frac{1}{3} - z_{1,1} - z_{1,3}} = 0$$

$$\Rightarrow \; z_{1,3}(\frac{1}{3} - z_{1,1} - z_{1,3}) + 3z_{1,1}(\frac{1}{3} - z_{1,1} - z_{1,3}) - 4z_{1,1}z_{1,3} = 0 \; \Rightarrow$$

$$-3z_{1,1}^2 + z_{1,1}(1 - 8z_{1,3}) + \frac{1}{3}z_{1,3} - z_{1,3}^2 = 0 \quad [4]$$

From [3] we obtain

$$-3z_{1,1}^2 + z_{1,1}(3 - 8z_{1,3}) + \frac{5}{3}z_{1,3} - z_{1,3}^2 = 0 \qquad [5]$$

$[4] - [5]$ gives

$$-2z_{1,1} - \frac{4}{3}z_{1,3} + \frac{1}{3} = 0 \ \Rightarrow z_{1,1} = \frac{1}{6} - \frac{2}{3}z_{1,3}$$

Substituting into [4] we obtain

$$-3(\frac{1}{6} - \frac{2}{3}z_{1,3}) + (\frac{1}{6} - \frac{2}{3}z_{1,3})(1 - 8z_{1,3}) + \frac{1}{3}z_{1,3} - z_{1,3}^2 = 0 \ \Rightarrow \ \frac{1}{12} - z_{1,3} + 3z_{1,3}^2 = 0$$

$\therefore \ z_{1,3} = \frac{1}{6}$ and $z_{1,1} = \frac{1}{18}$. But this is not a solution of equation [1] $\Box$

The following example illustrates that for priors $f^{(n)}$ continuous on $Int(V^{(n)})$ but not bounded on $V^{(n)}$ there can be relatively simple equality knowledge bases for which $f^{(n)}(\vec{x}|K)$ is undefined.

Example

$L^{(2)} = \{p, q\}$ and $K = \{E(p) = E(q)\}$

Then $V^{(2)}(K)$ can be parameterised as follows

$$< x_1, x_2, x_2, 1 - 2x_2 - x_1 >$$

where $0 < x_2 < \frac{1}{2}$ and $0 < x_1 < 1 - 2x_2$

Now if $f^{(2)} = d(\lambda, 2)$ then $f^{(2)}$ restricted to $V^{(n)}(K)$ is the proportional to the function

$$x_1^{\frac{\lambda}{2}-1} x_2^{\lambda-2} (1 - 2x_2 - x_1)^{\frac{\lambda}{2}-1}$$

for $0 < x_2 < \frac{1}{2}$ and $0 < x_1 < 1 - 2x_2$

Now clearly this function is not integrable on $V^{(2)}(K)$. In fact $f^{(2)} \restriction V^{(2)}(K)$ is not integrable on $V^{(2)}(K)$ for all $\lambda < 1$

For $\lambda > 1$ we have

$$CM_K^{d(\lambda,2)}(p \wedge \neg q) = \frac{\int_0^{\frac{1}{2}} \int_0^{1-2x_2} x_1^{\frac{\lambda}{2}-1} x_2^{\lambda-1} (1 - 2x_2 - x_1)^{\frac{\lambda}{2}-1} dx_1 dx_2}{\int_0^{\frac{1}{2}} \int_0^{1-2x_2} x_1^{\frac{\lambda}{2}-2} x_2^{\lambda-1} (1 - 2x_2 - x_1)^{\frac{\lambda}{2}-1} dx_1 dx_2}$$

$$= \frac{(\frac{1}{2})^{2\lambda-1} \beta(\lambda, \lambda)}{(\frac{1}{2})^{2\lambda-2} \beta(\lambda - 1, \lambda)} = \frac{1}{2} \frac{\lambda - 1}{2\lambda - 1}$$

Clearly then $\lim_{\lambda \searrow 1} CM_K^{d(\lambda,2)}(p \wedge \neg q) = \lim_{\lambda \searrow 1} CM_K^{d(\lambda,2)}(\neg p \wedge q) = 0$ which implies that $\lim_{\lambda \searrow 1} CM_K^{d(\lambda,2)}(p \leftrightarrow q) = 1$ where $\lambda \searrow 1$ means that λ tends to 1 from above.

Proposition 72 *(i) For any hierarchy of symmetric Dirichlet priors satisfying Marginality there exists $n > 0$ and K an equality knowledge base of $L^{(n)}$ such that $f^{(n)} \restriction V^{(n)}(K)$ is not integrable on $V^{(n)}(K)$.*

(ii) For any Dirichlet prior $d(\lambda, n)$ on $V^{(n)}$ such that $0 < \lambda < \frac{2^n}{2^n-1}(2^n - 2)$ there is an equality knowledge base of $L^{(n)}$ such that $d(\lambda, n) \restriction V^{(n)}(K)$ is not integrable on $V^{(n)}(K)$.

Proof

(i) By Marginality and proposition 25 $\{f^{(n)}\} = \{d(\lambda, n)\}$ where λ is independent of n

Let $n > \log_2(\lambda) + 1$ and $K = \{E(\alpha_1^{(n)}) = E(\alpha_2^{(n)})\}$. Then $V^{(n)}(K)$ can be parameterised as follows

$$< x_2, x_2, \ldots, x_{2^n-1}, 1 - 2x_2 - x_3, \ldots - x_{2^n-1} >$$

where $0 < x_2 < \frac{1}{2}$,$0 < x_3 < 1 - 2x_2,\ \ldots\ , 0 < x_{2^n-1} < 1 - 2x_2 - \ldots - x_{2^n-2}$

Hence, $d(\lambda, n) \restriction V^{(n)}(K)$ is proportional to the function

$$x_2^{\frac{\lambda}{2^{n-1}}-2} \prod_{i=3}^{2^n-1} x_i^{\frac{\lambda}{2^n}-1}(1 - 2x_2 - \sum_{i=3}^{2^n-1} x_i)^{\frac{\lambda}{2^n}-1}$$

for all $\vec{x} \in V^{(n)}(K)$

and $\lambda < 2^{n-1} \Rightarrow \frac{\lambda}{2^{n-1}} - 2 < -1$. Therefore, $d(\lambda, n) \restriction V^{(n)}(K)$ is not integrable on $V^{(n)}(K)$

(ii) Let $K = \{E(\alpha_1^{(n)}) = \ldots = E(\alpha_{2^n-1}^{(n)})\}$ then $V^{(n)}(K)$ can be parameterised as follows

$$< x_1, \ldots, x_1, 1 - (2^n - 1)x_1 > \quad 0 < x_1 < \frac{1}{2^n - 1}$$

Hence, $d(\lambda, n) \restriction V^{(n)}(K)$ is proportional to the function

$$x_1^{(2^n-1)\frac{\lambda}{2^n} - (2^n-1)}(1 - (2^n - 1)x_1)^{\frac{\lambda}{2^n}-1} \ 0 < x_1 < \frac{1}{2^n - 1}$$

for all $\vec{x} \in V^{(n)}(K)$

Now $\lambda < \frac{2^n}{2^n-1}(2^n - 2) \Rightarrow (2^n - 1)\frac{\lambda}{2^n} - (2^n - 1) < -1$ and therefore $d(\lambda, n) \restriction V^{(n)}(K)$ is not integrable on $V^{(n)}(K)$ $\square$

Clearly then there is no hierarchy of symmetric Dirichlet priors satisfying Marginality such that $\forall n > 0\ CM^{d(\lambda_n,n)}$ on equality knowledge bases is an equality inference process. However, any hierarchy of Dirichlet priors $\{d(\kappa 2^{n-1}, n)\}$ for $\kappa \geq 1$ is such that $\forall n > 0\ CM^{d(\kappa 2^{n-1},n)}$ on equality knowledge bases is an equality inference process and $\{CM^{d(\kappa 2^{n-1},n)}\}$ satisfies ZP Invariance

4.3 Bayesian Updating

In the following I relate the Centre of Mass inference process to Bayesian inference based on empirical data obtained from independent samples.

Let K be an open interval knowledge base and $freq_N$ be the information that for all $\theta \in SL^{(n)}$ $E(\theta)$ is interpreted as being the frequency of occurrence of the event $\underline{\theta} = 1$ in some independent sample of size N. That is

$$freq_N = \{E(\alpha_i^{(n)}) = \frac{r_i}{N} \mid r_i \in \{0, \ldots, N\}\ i = 1, \ldots, 2^n\}$$

The data base $K \cup freq_N$ might occur in practice as follows. Suppose an experts knowledge consists of the frequency of occurrence of events $\underline{\theta} = 1$ for $\theta \in SL^{(n)}$ in a large set of independent samples of size N. Now rather than combining this data to give the frequencies from one huge sample the expert might, for practical reasons, choose to use the information to make some more general inference regarding frequencies in samples of size N from this population. Given that any such inference would arise as a result of an expert observing patterns in the data it is not unreasonable to assume that it would be linear in nature.

Definition 73 *(i)* $V^{(n)}{}_N = \{\vec{x} \in V^{(n)} \mid x_i = \frac{r_i}{N}\ r_i \in \{0, \ldots, N\}\}$

(ii) $V^{(n)}{}_N(K) = V^{(n)}(K) \cap V^{(n)}{}_N$

Let $f^{(n)}(\vec{x})$ be the prior density on $V^{(n)}$ then from Bayes theorem we obtain the following posterior distribution

$$f^{(n)}(\vec{x}|K, freq_N) = \frac{L(K, freq_N|\vec{x})\ f^{(n)}(\vec{x})}{\int_{V^{(n)}} L(K, freq_N|\vec{x})\ f^{(n)}(\vec{x})dV^{(n)}}$$

where $L(K, freq_N|\vec{x})$ is the likelihood of the evidence $K \cup freq_N$ given $\vec{x}$ and

$$L(K, freq_N|\vec{x}) = \sum_{\vec{t} \in {V^{(n)}}_{N(K)}} C^N_{Nt_1 \dots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i}$$

therefore, for ${V^{(n)}}_N(K) \neq \emptyset$

$$f^{(n)}(\vec{x}|K, freq_N) = \frac{\sum_{\vec{t} \in {V^{(n)}}_{N(K)}} C^N_{Nt_1 \dots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x})}{\int_{V^{(n)}} \sum_{\vec{t} \in {V^{(n)}}_{N(K)}} C^N_{Nt_1 \dots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} f(\vec{x})dV^{(n)}}$$

Hence, the conditional expected value is given by

$$\mathcal{E}_{f^{(n)}}(x_j|K, freq_N) = \frac{\int_{V^{(n)}} x_j \sum_{\vec{t} \in {V^{(n)}}_{N(K)}} C^N_{Nt_1 \dots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x})dV^{(n)}}{\int_{V^{(n)}} \sum_{\vec{t} \in {V^{(n)}}_{N(K)}} C^N_{Nt_1 \dots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x})dV^{(n)}}$$

$$= \frac{\sum_{\vec{t} \in {V^{(n)}}_{N(K)}} C^N_{Nt_1 \dots Nt_{2^n-1}} \int_{V^{(n)}} x_j \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x})dV^{(n)}}{\sum_{\vec{t} \in {V^{(n)}}_{N(K)}} C^N_{Nt_1 \dots Nt_{2^n-1}} \int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x})dV^{(n)}}$$

Proposition 74 *If* $f^{(n)} \in \mathcal{C}$ *and* $f^{(n)}$ *is continuous on* $V^{(n)}$ *then for* $K \in IKB(L^{(n)})$

$$\lim_{N \to \infty} \mathcal{E}_{f^{(n)}}(x_j|K, freq_N) \ = \ CM_K^{f^{(n)}}(\alpha_j^{(n)})$$

Proof

Since

$$C^N_{Nt_1\ldots Nt_{2^n-1}} = \frac{1}{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} dV^{(n)}} \frac{N!}{(N+2^n-1)!}$$

we have, for ${V^{(n)}}_N(K) \neq \emptyset$

$$\mathcal{E}_{f^{(n)}}(x_j|K, freq_N) = \frac{\sum_{\vec{t}\in {V^{(n)}}_N(K)} \frac{\int_{V^{(n)}} x_j \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} dV^{(n)}}}{\sum_{\vec{t}\in {V^{(n)}}_N(K)} \frac{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} dV^{(n)}}}$$

Now if K is consistent then since $K \in IKB(L^{(n)})$ $\exists M \in \mathbb{N}$ such that $\forall N > M$ ${V^{(n)}}_N(K) \neq \emptyset$

Also as $N \to \infty$ $\prod_{i=1}^{2^n} x_i^{Nt_i}$ becomes increasingly peaked at its maximum point $< t_1 \ldots t_{2^n} >$. Since $f(\vec{x})$ is continuous and bounded on $V^{(n)}$ the same is true of $\prod_{i=1}^{2^n} x_i^{Nt_i} f(\vec{x})$. Therefore,

$$\lim_{N \to\infty} \mathcal{E}_{f^{(n)}}(x_j|K, freq_N) = \lim_{N \to\infty} \frac{\sum_{\vec{t}\in {V^{(n)}}_N(K)} t_i f^{(n)}(\vec{t})}{\sum_{\vec{t}\in {V^{(n)}}_N(K)} f^{(n)}(\vec{t})} =$$

$$\frac{\int_{V^{(n)}(K)} x_j f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}(K)} f^{(n)}(\vec{x}) dV^{(n)}}$$

since $f^{(n)}$ is continuous on $V^{(n)}(K)$ $\square$

This result illustrates that if $E(\theta)$ is interpreted as being the frequency of occurrence of the event $\underline{\theta} = 1$ in a sequence of independent trials and K, an open interval probability state, is interpreted as knowledge regarding such frequencies then in the limit as the number trials tends to infinity the expected value of x_j relative to prior densities continuous on $V^{(n)}$ and calculated according to Bayes theorem agrees with $CM_K^{f^{(n)}}(\alpha_j^{(n)})$.

In addition, we consider the following alternative interpretation. If $\mathcal{F}$ is a distribution on $\mathcal{V}^{(n)}$ then by the de Finetti Representation theorem (see [10]) the probability of the vector of frequencies of events $\underline{\alpha_i^{(n)}} = 1$ being $\vec{t} \in {V^{(n)}}_N$ is

$$P_N(\vec{t}) = \frac{\int_{V^{(n)}} C^N_{Nt_1 \ldots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} d\mathcal{F}}{\sum_{\vec{t} \in {V^{(n)}}_N} \int_{V^{(n)}} C^N_{Nt_1 \ldots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} d\mathcal{F}}$$

So that for the case discussed above

$$P_N(\vec{t}) = \frac{\int_{V^{(n)}} C^N_{Nt_1 \ldots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}{\sum_{\vec{t} \in {V^{(n)}}_N} \int_{V^{(n)}} C^N_{Nt_1 \ldots Nt_{2^n-1}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}$$

Then we have, for ${V^{(n)}}_N(K) \neq \emptyset$

$$\mathcal{E}_{P_N}(t_j|K) = \frac{\sum_{\vec{t} \in {V^{(n)}}_{N(K)}} t_j P_N(\vec{t})}{\sum_{\vec{t} \in {V^{(n)}}_{N(K)}} P_N(\vec{t})}$$

Proposition 75 *If* $f^{(n)} \in \mathcal{C}$ *and* $f^{(n)}$ *is continuous on* $V^{(n)}$ *then for* $K \in IKB(L^{(n)})$

$$\lim_{N \to \infty} \mathcal{E}_{P_N}(t_j|K) = CM_K^{f^{(n)}}(\alpha_j^{(n)})$$

Proof

As in the proof of Proposition 74 since $K \in IKB(L^{(n)})$ there exists $M \in \mathbb{N}$ such that $\forall N > M$ ${V^{(n)}}_N(K) \neq \emptyset \therefore$

$$\mathcal{E}_{P_N}(t_j|K) = \frac{\sum_{\vec{t} \in {V^{(n)}}_{N(K)}} t_j \frac{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} dV^{(n)}}}{\sum_{\vec{t} \in {V^{(n)}}_{N(K)}} \frac{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} dV^{(n)}}}$$

Then by the properties of $\prod_{i=1}^{2^n} x_i^{Nt_i}$ as N tends to infinity described above in the proof of proposition 74 we have since $f^{(n)}$ is continuous on $V^{(n)}$

$$\lim_{N\to\infty} \mathcal{E}_{P_N}(t_j|K) = \lim_{N\to\infty} \frac{\sum_{\vec{t}\in {V^{(n)}}_{N(K)}} t_j f^{(n)}(\vec{t})}{\sum_{\vec{t}\in {V^{(n)}}_{N(K)}} f^{(n)}(\vec{t})}$$

$$= CM_K^{f^{(n)}}(\alpha_j^{(n)})$$

□

Hence, we can also think of $CM_K^{f^{(n)}}(\alpha_j^{(n)})$ as the limit as N tends to infinity of the expected frequency of the event $\underline{\alpha_j^{(n)}} = 1$ in some sequence of independent trials relative to a measure P_N on ${V^{(n)}}_N$ determined by $f^{(n)}$

Suppose that K is a knowledge base specifying the frequency of occurrence of the event $\underline{\alpha_i^{(n)}} = 1$ as t_i for $i = 1, \ldots, 2^n$ in an independent sample of size N then

$$\mathcal{E}_{f^{(n)}}(x_j|K) = \frac{\int_{V^{(n)}} x_j \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}{\int_{V^{(n)}} \prod_{i=1}^{2^n} x_i^{Nt_i} f^{(n)}(\vec{x}) dV^{(n)}}$$

Now for $f^{(n)} = d(\lambda, n)$ we obtain

$$\mathcal{E}_{f^{(n)}}(x_j|K) = \frac{Nt_j + \frac{\lambda}{2^n}}{N + \lambda}$$

In fact, rather more is true and $\mathcal{E}_{f^{(n)}}(x_j|K) = \frac{Nt_j + \frac{\lambda}{2^n}}{N+\lambda}$ if and only if $f^{(n)} = d(\lambda, n)$

This continuum of inductive inference processes, often referred to as Carnap's continuum, was first proposed by Johnson [15] and later by Carnap [5] as being that which was characterised by a number of symmetry principles on a confirmation function, quantifying the degree to which a hypothesis is confirmed by a body of evidence.

It is interesting to note that an original aim of Carnaps programme was to justify the choice of a unique confirmation function for a given language in terms of logical axioms. In the context of Carnaps continuum this is equivalent to the problem of justifying the choice of a particular value of λ. In the light of Carnaps failure to suggest such axioms a pertinent question is whether, in the context of Inexact Reasoning, there are principles characterising a unique hierarchy of Dirichlet priors?

In the next chapter we show that some of the axioms on priors described above together with axioms relating to independence of the random propositional variables do characterise a unique hierarchy of Dirichlet priors. We also put forward an argument for the choice of a different hierarchy of priors based on maximising the *a priori* expected level of dependency between random propositional variables.

Chapter 5

Independence

5.1 Independence of Random Propositional Variables

In the previous chapters we have only considered priors conditionalised on knowledge bases which consist of linear constraints on a probability function E. However, a common assumption in statistics is that various events are independent. In the context of Inexact Reasoning this would correspond to conditionalising on the knowledge that the random propositional variables of the language are independent. In this chapter we give a definition of the conditional density relative to this non linear knowledge base and then propose a number axioms relating to this density.

Let $Ind_n = \{E(\bigwedge_{i=1}^n p_i^{\epsilon_i^j}) = \prod_{i=1}^n E(p_i^{\epsilon_i^j})\ ,\ j = 1, \ldots 2^n\}$ where ϵ_i^j is the i'th coefficient of the binary expansion of $2^n - j$. So Ind_n is the knowledge that the random propositional variables $\underline{p_i}$ are independent for $i = 1, \ldots, n$.

Definition 76

$$V^{(n)}{}_{Ind} = \{\vec{x} \in V^{(n)} | x_j = \prod_{i=1}^{n} z_i^{\epsilon_i^j} \ , \ j = 1, \ldots 2^n\}$$

where $z_j = \sum_{i=1}^{2^{n-j}} \sum_{k=1}^{2^{j-1}} x_{2^j i - k}$ *and* $z_j^1 = z_j$, $z_j^0 = (1 - z_j)$

Note that $V^{(n)}{}_{Ind}$ *is not convex.*

Clearly, every point in $V^{(n)}{}_{Ind}$ uniquely defines a probability function satisfying Ind_n. In addition, there is a natural correspondence between points in $\{\vec{z} \in \mathbb{R}^n \ | z_i \in [0,1] \ , \ i = 1, \ldots, n\}$ and points in $V^{(n)}{}_{Ind}$ where $< z_1, \ldots, z_n >$ is associated with the $< \prod_{i=1}^{n} z_i^{\epsilon_i^1}, \ldots, \prod_{i=1}^{n} z_i^{\epsilon_i^{2^n}} >$ and vica versa. That is to say , in the presence of the independence constraints Ind_n the values of $E(p_i)$ for $i = 1, \ldots, n$ uniquely determines a probability function on $SL^{(n)}$. Hence, in keeping with the conditionalisation assumption, we are motivated to make the following definition.

Definition 77 *Let* Z_i *denote the random variable* $E(p_i)$ *for* $i = 1, \ldots, n$. *Then if* $f^{(n)} \in \mathcal{C}$ *satisfying Strong Non Nullity we define the following density on* Z_i *for* $i = 1, \ldots, n$ *relative to* Ind_n

$$f^{(n)}{}_{Ind}(\vec{z}) = \frac{f^{(n)}(\prod_{i=1}^{n} z_i^{\epsilon_i^1}, \ldots, \prod_{i=1}^{n} z_i^{\epsilon_i^{2^n}})}{\int_0^1 \ldots \int_0^1 f^{(n)}(\prod_{i=1}^{n} z_i^{\epsilon_i^1}, \ldots, \prod_{i=1}^{n} z_i^{\epsilon_i^{2^n}}) d\vec{z}}$$

if $\int_0^1 \ldots \int_0^1 f^{(n)}(\prod_{i=1}^{n} z_i^{\epsilon_i^1}, \ldots, \prod_{i=1}^{n} z_i^{\epsilon_i^{2^n}}) d\vec{z}$ *exists and is left undefined otherwise*

A11 : Total Independence

Let $f^{(n)} \in \mathcal{C}$ be a prior density on $V^{(n)}$ satisfying Strong Non Nullity and such that ${f^{(n)}}_{Ind}$ is defined then $f^{(n)}$ satisfies Total Independence if ${f^{(n)}}_{Ind}(\vec{z}) = \prod_{i=1}^{n} h_i^{(n)}(z_i)$ where

$$h_i^{(n)}(z_i) = \int_0^1 \dots \int_0^1 {f^{(n)}}_{Ind}(\vec{z}) dz_n \dots dz_{i-1} dz_{i+1} \dots dz_1$$

is the marginal density of the random variable Z_i

In other words , the random variables $E(p_i)$ for $i = 1, \dots, n$ are independent given Ind_n.

This axiom is motivated by the idea that since the random propositional variables $\underline{p_1}, \dots, \underline{p_n}$ are independent it should not be possible to infer anything regarding the value of $E(p_i)$ for some $i \in \{1, \dots, 2^n\}$ from knowledge of $E(p_j)$ for $j \neq i$. Thus, independence of the random propositional variables p_i for $i = 1, \dots, n$ should imply independence of the random variables $E(p_i)$ for $i = 1, \dots, n$.

A12 : Marginality under Independence

Let $\{f^{(n)}\}$ be a hierarchy of priors such that $\forall n > 0$ ${f^{(n)}}_{Ind}$ is defined $f^{(n)} \in \mathcal{C}$ and satisfies Strong Non Nullity and Weak Renaming then $\{f^{(n)}\}$ satisfies Marginality under Independence if

$$\int_0^1 \dots \int_0^1 {f^{(n)}}_{Ind}(\vec{z}) dz_n \dots dz_{i+1} dz_{i-1} \dots dz_1 = f^{(1)}(z_i)$$

Less formally this states that the problem of giving a value to $E(p_i)$ in the context of $n-1$ other propositional variables about which all that is known is that they are independent of each other and of p_i is equivalent to the problem of giving a value to $E(p_i)$ when $\underline{p_i}$ is the only random propositional variable under consideration.

Clearly then we have for any hierarchy of priors $\{f^{(n)}\}$ such that $\forall n > 0$ $f^{(n)} \in \mathcal{C}$ and satisfies Strong Non Nullity, Weak Renaming and Total Independence that if $\{f^{(n)}\}$ satisfies Marginality under Independence then $\forall n > 0$

$$f^{(n)}{}_{Ind}(\vec{z}) = \frac{f^{(n)}(\prod_{i=1}^{n} z_i^{\epsilon_i^1}, \ldots, \prod_{i=1}^{n} z_i^{\epsilon_i^{2^n}})}{\int_0^1 \cdots \int_0^1 f^{(n)}(\prod_{i=1}^{n} z_i^{\epsilon_i^1}, \ldots, \prod_{i=1}^{n} z_i^{\epsilon_i^{2^n}}) d\vec{z}}$$

$$= \prod_{i=1}^{n} f^{(1)}(z_i)$$

Proposition 78 *Let $\{d(\lambda_n, n)\}$ be a hierarchy of symmetric Dirichlet priors such that $\forall n > 0$ $f^{(n)}{}_{Ind}$ is defined. Then $\{d(\lambda_n, n)\}$ satisfies Marginality under Independence iff $\forall n > 0$ $\lambda_n = \lambda_1 + 2^n - 2$*

Proof

Firstly note that any hierarchy of symmetric Dirichlet priors $\{d(\lambda_n, n)\}$ $f^{(n)}{}_{Ind}$ is defined if and only if $\lambda_n > 2^n - 2$. Furthermore, any such hierarchy satisfying this condition satisfies Total Independence since

$$h^{(n)}(z_1) = \frac{\int_0^1 \cdots \int_0^1 \prod_{i=1}^{n} z_i^{2^{n-1}(\frac{\lambda_n}{2^n}-1)} (1-z_i)^{2^{n-1}(\frac{\lambda_n}{2^n}-1)} dz_n \ldots dz_2}{\int_0^1 \cdots \int_0^1 \prod_{i=1}^{n} z_i^{2^{n-1}(\frac{\lambda_n}{2^n}-1)} (1-z_i)^{2^{n-1}(\frac{\lambda_n}{2^n}-1)} dz_n \ldots dz_1}$$

$$= (\beta(\frac{\lambda_n - 2^n + 2}{2}, \frac{\lambda_n - 2^n + 2}{2}))^{-1} z_1^{2^{n-1}(\frac{\lambda_n}{2^n}-1)} (1-z_1)^{2^{n-1}(\frac{\lambda_n}{2^n}-1)}$$

and

$$f^{(n)}{}_{Ind}(\vec{z}) = (\beta(\frac{\lambda_n - 2^n + 2}{2}, \frac{\lambda_n - 2^n + 2}{2}))^{-n} \prod_{i=1}^{n} z_i^{2^{n-1}(\frac{\lambda_n}{2^n}-1)}(1-z_i)^{2^{n-1}(\frac{\lambda_n}{2^n}-1)}$$

Now by Marginality under Independence

$$(\beta(\frac{\lambda_n - 2^n + 2}{2}, \frac{\lambda_n - 2^n + 2}{2}))^{-1} z^{2^{n-1}(\frac{\lambda_n}{2^n}-1)}(1-z)^{2^{n-1}(\frac{\lambda_n}{2^n}-1)}$$

$$= (\beta(\frac{\lambda_1}{2}, \frac{\lambda_1}{2}))^{-1} z^{\frac{\lambda_1}{2}-1}(1-z)^{\frac{\lambda_1}{2}-1} \Rightarrow \lambda_n = \lambda_1 + 2^n - 2$$

□

Intuitively one might expect Marginality under Independence to be a weakening of Marginality. However, by the above result, it can be seen that there is no hierarchy of Dirichlet priors satisfying both principles.

Corollary 79 *Let $\{f^{(n)}\}$ be a hierarchy of priors such that $\forall n > 0$ $f^{(n)}{}_{Ind}$ is defined, $f^{(n)} \in \mathcal{C}_M^2$ and satisfies Weak Renaming, Strong Non Nullity and Relative Ignorance then given the conditionalisation assumption $\{f^{(n)}\}$ satisfies Marginality under Independence and Total Improbability iff $\forall n > 0$ $f^{(n)} = d(2^n, n)$ which is the uniform prior on $V^{(n)}$*

Proof

By Weak Renaming , Relative Ignorance and Strong Non Nullity we have by theorem 47 that $\forall n > 0$ $f^{(n)} = d(\lambda_n, n)$

By Marginality under Independence we have $\lambda_n = \lambda_1 + 2^n - 2$

However, by Total Improbability and corollary 52 we have $\lambda_n = 2^n\kappa$ for $\kappa \in \mathbb{R}^{>0}$

Hence $\lambda_1 = 2\kappa$ and $\lambda_n = \kappa 2 + 2^n - 2 = \kappa 2^n \;\Rightarrow\; \kappa = 1$ $\square$

Thus the axioms A12, A10, A2, A5, A7, in the presence of the above smoothness assumptions, provide a possible epistemological justification for the choice of the hierarchy of uniform priors.

In the above we have considered priors conditionalised on the knowledge that the random propositional variables are independent. In the sequel we discuss some *a priori* assumptions regarding independence of the random propositional variables associated with the choice of particular hierarchies of Dirichlet priors.

5.2 Covariance between Random Propositional variables

A natural question is what is the average level of dependence between random propositional variables with distributions generated according to some Dirichlet prior $d(\lambda, n)$. In other words, what is the average covariance relative to some symmetric Dirichlet prior $d(\lambda, n)$ between the random propositional variables $\underline{p_i}$ and $\underline{p_j}$ for $j \neq i$ and $i, j \in \{1, \ldots, n\}$.

Now

$$\mathcal{E}_{d(\lambda,n)}(Cov(\underline{p_i}, \underline{p_j})) = \mathcal{E}_{d(\lambda,n)}((E(p_i \wedge p_j) - E(p_i)E(p_j))^2)$$

Then w.l.o.g consider

$$\mathcal{E}_{d(\lambda,n)}((E(p_n \wedge p_{n-1}) - E(p_n)E(p_{n-1}))^2) = \mathcal{E}_{d(\lambda,n)}((X - (X+Y)(X+W))^2)$$

where $X = \sum_{i=1}^{\frac{J}{4}} x_{4i-3} = E(p_n \wedge p_{n-1})$, $Y = \sum_{i=1}^{\frac{J}{4}} x_{4i-2} = E(p_n \wedge \neg p_{n-1})$ and $W = \sum_{i=1}^{\frac{J}{4}} x_{4i-1} = E(\neg p_n \wedge p_{n-1})$ for $J = 2^n$

$$= \mathcal{E}_{d(\lambda,n)}((X - X^2 - YX - WX - WY)^2) = \mathcal{E}_{d(\lambda,n)}((X(1 - X - Y - W) - WY)^2)$$

$$= \mathcal{E}_{d(\lambda,n)}((XZ - WY)^2)$$

where $Z = 1 - X - Y - W = E(\neg p_n \wedge \neg p_{n-1}) = \sum_{i=1}^{\frac{J}{4}} x_{4i}$

$$= \mathcal{E}_{d(\lambda,n)}(X^2Z^2 - 2XZWY + W^2Y^2) = 2\mathcal{E}_{d(\lambda,n)}(X^2Z^2) - 2\mathcal{E}_{d(\lambda,n)}(XZWY)$$

by symmetry of $d(\lambda, n)$

Now again by the symmetry of $d(\lambda, n)$

$$\mathcal{E}_{d(\lambda,n)}(X^2Z^2) = (\frac{J}{4})^2\mathcal{E}_{d(\lambda,n)}(x_i^2x_j^2) + (\frac{J}{4})^2(\frac{J}{4} - 1)^2\mathcal{E}_{d(\lambda,n)}(x_ix_jx_kx_l)$$
$$+2(\frac{J}{4})^2(\frac{J}{4} - 1)\mathcal{E}_{d(\lambda,n)}(x_i^2x_jx_k)$$

$$\mathcal{E}_{d(\lambda,n)}(XZWY) = (\frac{J}{4})^2\mathcal{E}_{d(\lambda,n)}(x_ix_jx_kx_l)$$

where $i, j, k, l \in \{1, \ldots, 2^n\}$ are distinct

Further

$$\mathcal{E}_{d(\lambda,n)}(x_i^2x_j^2) = \frac{(\frac{\lambda}{J} + 1)^2(\frac{\lambda}{J})^2}{\lambda(\lambda+1)(\lambda+2)(\lambda+3)}$$

$$\mathcal{E}_{d(\lambda,n)}(x_i^2x_jx_k) = \frac{(\frac{\lambda}{J})^3(\frac{\lambda}{J} + 1)}{\lambda(\lambda+1)(\lambda+2)(\lambda+3)}$$

$$\mathcal{E}_{d(\lambda,n)}(x_i x_j x_k x_l) = \frac{(\frac{\lambda}{J})^4}{\lambda(\lambda+1)(\lambda+2)(\lambda+3)}$$

$$\therefore\ \mathcal{E}_{d(\lambda,n)}((E(p_n \wedge p_{n-1}) - E(p_n)E(p_{n-1}))^2) =$$

$$\frac{\frac{J^2}{16}(\frac{\lambda}{J}+1)^2(\frac{\lambda}{J})^2 + 2\frac{J^2}{16}(\frac{J}{4}-1)(\frac{\lambda}{J})^3(\frac{\lambda}{J}+1) + (\frac{J^2}{16}(\frac{J}{4}-1)^2 - \frac{J^4}{4^4})(\frac{\lambda}{J})^4}{\lambda(\lambda+1)(\lambda+2)(\lambda+3)}$$

$$= \frac{\lambda}{16(\lambda+1)(\lambda+3)}$$

Now it can be seen that for any hierarchy of Dirichlet priors $\{d(\kappa 2^n, n)\}$, $\kappa \in \mathbb{R}^{>0}$ satisfying Total Improbability $\lim_{n\to\infty} \mathcal{E}_{d(\lambda,n)}(Cov(\underline{p_i}, \underline{p_j})) = 0$

In addition, we have

$$\mathcal{E}_{d(\lambda,n)}((E(p_i) - \frac{1}{2})^2) = \frac{1}{4(\lambda+1)}$$

which implies that if $\{d(\lambda_n, n)\}$ satisfies Total Improbability then $\lim_{n\to\infty} \mathcal{E}_{d(\lambda,n)}((E(p_i) - \frac{1}{2})^2) = 0$

In other words, for all except small values of n, probability functions generated according to Dirichlet priors from a hierarchy satisfying Total Improbability will on average have the properties that the propositional variables are almost independent and the expected value of propositional variables are close to $\frac{1}{2}$. Now it could be argued that this is not representative of the sort of distributions for which, in practice, we would wish to make inductive inferences and construct expert systems.

In practical situations the random propositional variables tend to be quite highly correlated and have a much higher spred of expected values.

Furthermore, it does not seem desirable for assumptions of independence to be *a priori* part of the model. Rather, such information should be included in terms of a conditional prior as suggested in the previous section.

If, however, we consider hierarchies of Dirichlet priors satisfying Marginality clearly $\mathcal{E}_{d(\lambda,n)}(Cov(\underline{p_i},\underline{p_j}))$ and $\mathcal{E}_{d(\lambda,n)}((E(p_i) - \frac{1}{2})^2)$ are constant for all $n > 0$. Therefore, in view of the above comments it might be justifiable to choose as a 'natural' hierarchy of Dirichlet priors, that which satisfies Marginality and for which the value of $\mathcal{E}_{d(\lambda,n)}(Cov(\underline{p_i},\underline{p_j}))$ is maximal. It can easily be seem that the function $\frac{\lambda}{16(\lambda+1)(\lambda+3)}$ has a unique maximum point at $\lambda = \sqrt{3}$ and hence the hierarchy of priors with above properties is $\{d(\sqrt{3},n)\}$ for which we have that

$$\mathcal{E}_{d(\sqrt{3},n)}(Cov(\underline{p_i},\underline{p_j})) \approx 8.3734 \times 10^{-3}$$

and

$$\mathcal{E}_{d(\sqrt{3},n)}((E(p_i) - \frac{1}{2})^2) \approx 0.0915$$

Now the variance of $E(p_i \wedge p_j)$ with respect to $d(\lambda, n)$ is given by

$$\mathcal{E}_{d(\lambda,n)}(E(p_i \wedge p_j) - \frac{1}{4})^2 =$$

$$\mathcal{E}_{d(\lambda,n)}(E(p_i \wedge p_j)^2) - \frac{1}{16}$$

$$= \frac{J}{4}\mathcal{E}_{d(\lambda,n)}(x_i^2) + \frac{J}{4}(\frac{J}{4} - 1)\mathcal{E}_{d(\lambda,n)}(x_i x_j) - \frac{1}{16}$$

$$\frac{\frac{J}{4}\frac{\lambda}{J}(\frac{\lambda}{J} + 1) + \frac{J}{4}(\frac{J}{4} - 1)(\frac{\lambda}{J})^2}{\lambda(\lambda + 1)} - \frac{1}{16}$$

$$\frac{3}{16(\lambda + 1)}$$

Hence, the average degree to which $E(p_i \wedge p_j)$ differs from $E(p_i)E(p_j)$ relative to the total variance of $E(p_i \wedge p_j)$ is

$$\sqrt{\frac{\mathcal{E}_{d(\lambda,n)}(Cov(\underline{p_i}, \underline{p_j}))}{\mathcal{E}_{d(\lambda,n)}((E(p_i \wedge p_j) - \frac{1}{4})^2)}} = \sqrt{\frac{\lambda}{3(\lambda + 3)}}$$

Again this is maximal at $\lambda = \sqrt{3}$ for which the value is approximately 0.35

Even if we accept the argument that $\mathcal{E}_{d(\lambda,n)}(Cov(\underline{p_i}, \underline{p_j}))$ should be large for all $n > 0$ it remains unclear as to exactly how large it should be. In the case of $d(\sqrt{3}, n)$ we would suggest that the relative value of 0.35 is of sufficient magnitude to ensure that distributions generated according to $d(\sqrt{3}, n)$ will not tend to treat the random propositional variables of $L^{(n)}$ as being close to independent.

Bibliography

[1] J.Aczel Lectures on Functional Equations and their Applications
Academic Press (1966)

[2] T.M.Apostal Mathematical Analysis
Addison-Wesley Publishing Company (1974)

[3] G.De Barra Introduction to Measure Theory
Van Nostrand Reinhold Company (1974)

[4] C.W.Burril Measure, Integration, and Probability
McGraw-Hill Inc (1972)

[5] R.Carnap. Continuum of Inductive methods
University of Chicago Press, Chicago (1952)

[6] H.G Eggleston Convexity
Cambridge University Press (1958)

[7] H.B.Enderton A Mathematical Introduction to Logic
Academic Press (1972)

[8] W.Feller An Introduction to Probability Theory and its Applications Vol2
John Wiley and Sons, Inc (1966)

[9] T.L.Fine Theories of Probabilities: an examination of foundations Academic Press (1973)

[10] B.De Finetti Foresight: its logical laws, its subjective sources.
Translated from the french of 1937 by H.Kyburg in Studies in Subjective Probability (H.E Kyburg and H.E Smokler eds) p95 - p158 Wiley (1980)

[11] I.J.Good The Estimation of Probabilities: An Essay on Modern Bayesian Methods
M.I.T Press (1965)

[12] I.J.Good How to Estimate Probabilities
Journal of the Institute of Mathematics and its Application (1966)

[13] I.J.Good A Bayesian Significance Test for Multinomial Distributions
Journal of the Royal Statistical Society series B 29 p 399-431 (1967)

[14] I.J.Good Good Thinking: The Foundations of Probability and its Applications
University of Minnesota Press (1983)

[15] W.E.Johnson. Probability: The Deductive and Inductive Problems,
Mind, Vol. XLI, No. 164, p27.(1932)

[16] W.Kecs The Convolution Product: and some applications
Editura Academiei (1982)

[17] E.Mendelson Introduction to Mathematical Logic
Wadsworth and Brooks/Cole Advanced Books and Software (1987)

[18] J.Paris, A.Vencovska A method for Updating Justifying Minimum Cross Entropy

To appear in The Journal of Approximate Reasoning.

[19] J.Paris, A.Vencovska Principles of Uncertain Reasoning

Proceedings of the second International Colloquium on Cognitive Science, San Sebastian, Spain (1991)

[20] J.Paris, A.Vencovska A Method for Updating that Justifies Minimum Cross Entropy

International Journal of Approximate Reasoning (1992)

[21] J.Paris, A.Vencovska, G.M.Wilmers A Note on Objective Inductive Inference

Proceedings of the First World Conference on the Foundations of Artificial Intelligence, Ed. M de Glas and D.Gabbay, Association Francaise pour l'Intelligence Artificialle, Paris , pp407-412. (1991)

[22] J.Paris The Uncertain Reasoner's Companion - a Mathematical Perspective

Cambridge University Press (1994)

[23] S.L.Zabel Symmetry and its Discontents.

Causation, Chance, and Credence Vol 1, pp155-190

[24] E.Watson Laplace Transforms and Applications

Van Nostrand Reinhold Company (1981)

[25] G.M.Wilmers, A.Vencovska, J.Paris

A Natural Prior Probability Distribution Derived From The Propositional Calculus

To appear in The Annals of Pure and Applied Logic

[26] S.Wilks Mathematical Statistics

John Wiley and Sons,Inc (1944)