ARTIFICIAL INTELLIGENCE FRONTIERS IN STATISTICS

AI AND STATISTICS III

Edited by
D.J. Hand

CHAPMAN & HALL

Contents

List of contributors ix

Introduction D.J. Hand xv

PART ONE Statistical expert systems 1

1 DEXPERT: an expert system for the design of experiments 3
T.J. Lorenzen, L.T. Truss, W.S. Spangler, W.T. Corpus and A.B. Parker

2 Inside two commercially available statistical expert systems 17
J.F.M. Raes

3 AMIA: Aide à la Modélisation par l'Intelligence Artificielle (expert system for simulation modelling and sectoral forecasting) 31
M. Ollivier, R. Arrus, M.-A. Durillon, S. Robert and B. Debord

4 An architecture for knowledge-based statistical support systems 39
A. Prat, E. Edmonds, J.M. Catot, J. Lores, J. Galmes and P. Fletcher

5 Enhancing explanation capabilities of statistical expert systems through hypertext 46
P. Hietala

6 Measurement scales as metadata 54
D.J. Hand

PART TWO Belief networks 65

7 On the design of belief networks for knowledge-based systems 67
B. Abramson

8 Lack-of-information based control in graphical belief systems 82

9 Adaptive importance sampling for Bayesian networks applied to filtering problems 90
A.R. Runnalls

10 Intelligent arc addition, belief propagation and utilization of parallel processors by probabilistic inference engines 106
A. Ranjbar and M. McLeish

11 A new method for representing and solving Bayesian decision problems 109
P.P. Shenoy

PART THREE Learning 139

12 Inferring causal structure in mixed populations 141
C. Glymour, P. Spirtes and R. Scheines

13 A knowledge acquisition inductive system guided by empirical interpretation of derived results 156
K. Tsujino and S. Nishida

14 Incorporating statistical techniques into empirical symbolic learning systems 168
F. Esposito, D. Malerba and G. Semeraro

15 Learning classification trees 182
W. Buntine

16 An analysis of two probabilistic model induction techniques 202
S.L. Crawford and M. Fung

PART FOUR Neural networks 215

17 A robust back propagation algorithm for function approximation 217
D.S. Chen and R.C. Jain

18 Maximum likelihood training of neural networks 241
H. Gish

19 A connectionist knowledge acquisition tool: CONKAT 256
A. Ultsch, R. Mantyk and G. Halmans

20 Connectionist, rule-based, and Bayesian decision aids: an empirical comparison 264
S. Schwartz, J. Wiles, I. Gough and S. Phillips

PART FIVE Text manipulation 279

21 Statistical approaches to aligning sentences and identifying word correspondences in parallel texts: a report on work in progress 281
W.A. Gale and K.W. Church

22 Probabilistic text understanding 295
R.P. Goldman and E. Charniak

23 The application of machine learning techniques in subject classification 312
I. Kavanagh, C. Ward and J. Dunnion

PART SIX Other areas 325

24 A statistical semantics for causation 327
J. Pearl and T.S. Verma

25 Admissible stochastic complexity models for classification problems 335
P. Smyth

26 Combining the probability judgements of experts: statistical and artificial intelligence approaches 348
L.A. Cox

27 Randomness and independence in non-monotonic reasoning 362
E. Neufeld

28 Consistent regions in probabilistic logic when using different norms 370
D. Bouchaffra

29 A decision theoretic approach to controlling the cost of planning 387
L. Hartman

Index 401

Consistent regions in probabilistic logic when using different norms

28

D. Bouchaffra

INTRODUCTION

Ordinary logic is very often used when dealing with certainty, but in many fields when we have to treat uncertainty or 'fuzziness' boolean logic appears a very poor concept. Several artificial intelligence (AI) or linguistic applications require the ability to reason with uncertain information. The semantic generalization of logic in which the truth values of sentences are probability values between 0 and 1 seems to be necessary, assignment of probability values could be consistent or inconsistent.

A lot of mathematical work on probabilistic inference already exists (Suppes, 1966; Zadeh, 1975). For instance, the MYCIN expert system handled uncertain knowledge, and the PROSPECTOR system used a method based on Bayes's rule (Duda *et al.*, 1976). In this chapter, our interest is in providing a new way of extending Nilsson's (1986) probabilistic logic by exploring different metrics. We adopt a geometric approach and discuss the relative notion of an extreme vector.

The second section defines the notion of consistency and the probability concept. The third section presents the algebraic and geometric interpretation. The next four sections present several ways of obtaining the consistent region according to metrics. We explore the infinite norm, the L_1 norm, the Euclidean norm, and also their corresponding isometries. The interpretation of Bayes's theorem when using the L_1 norm is discussed in the eighth section.

CONSISTENCY AND PROBABILITY

We define these concepts by an example. Let us consider the set of sentences $\mathcal{S} = \{P, \neg P \vee Q, Q\}$, so that we expect to obtain among all possible worlds (2^3) just 2^2 consistent possible worlds; other possible worlds are inconsistent.

Many methods exist for determining the sets of consistent truth values, given a set of sentences. A basic method is that called the **binary semantic tree** (Kleene, 1987).

Artificial Intelligence Frontiers in Statistics: AI and statistics III. Edited by D.J. Hand. Published in 1993 by Chapman & Hall, London. ISBN 0 412 40710 8

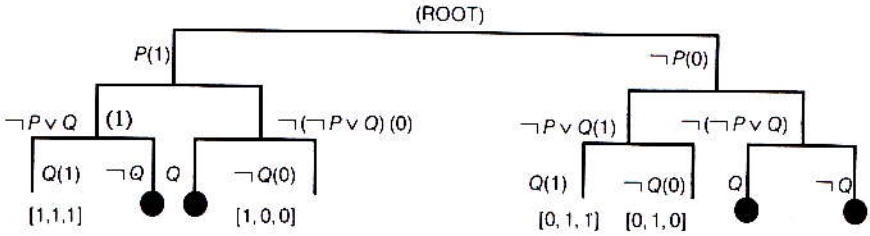

Figure 28.1 The binary tree.

We create nodes by branching left or right, depending on whether or not we assign one sentence a value of TRUE or FALSE respectively.

At each node, we branch left or right, depending on whether or not we assign one of the sentences in $\mathscr{S}$ a value of TRUE or FALSE, respectively. Just below the root, we branch the truth value of the next sentences of our set $\mathscr{S}$, and the procedure continues until the last sentence of $\mathscr{S}$. We obtain at the end a set of paths corresponding to unique assignments and we close off those paths corresponding to inconsistent valuations, as illustrated in Fig. 28.1. The black circles correspond to the closed paths (inconsistent evaluation), the numbers '1' and '0' at the right of each sentence correspond, respectively, to the evaluation TRUE and FALSE of the sentence. Thus we can write the matrix corresponding to the diagram, which consistent matrix we will denote $\boldsymbol{C}$:

$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

We can choose the probability of any sentence S as $P(S) = \sum_i P(\mathbf{V}_i)^* \psi_{\mathbf{V}_i}(S)$ where $P(\mathbf{V}_i)$ is the probability that the actual world $\mathbf{V}_a$ is equal to $\mathbf{V}_i$, the characteristic function ψ is defined as:

$$\psi_{\mathbf{V}_i}(S) = \begin{cases} 1 & \text{if } S \text{ is true in } \mathbf{V}_i \\ 0 & \text{if } S \text{ is false in } \mathbf{V}_i \end{cases}$$

$\mathbf{V}_i$ are the column vectors of the matrix $\boldsymbol{C}$, and i belongs to the set $\{1, 2, 3, 4\}$. This definition of the probability associated with a sentence gives a mathematical and geometric interpretation of the consistent area in the space of the three sentences.

THE ALGEBRAIC AND GEOMETRIC INTERPRETATION

Let us consider the following mapping between the space of possible worlds and the one of sentences. This mapping is the homomorphism associated with the consistent matrix $\boldsymbol{C}$ called $\boldsymbol{\Psi}$. Let X and Y denote, respectively, the space of possible words and that of sentences. We can plot the following diagram corresponding to our example:

$$X \longrightarrow \Psi \longrightarrow Y$$

$$\mathbf{V} = \begin{bmatrix} P(\mathbf{V}_1) \\ P(\mathbf{V}_2) \\ P(\mathbf{V}_3) \\ P(\mathbf{V}_4) \end{bmatrix} \dashrightarrow \Psi(\mathbf{V}) = \Pi = \begin{bmatrix} P(P) \\ P(\neg P \vee Q) \\ P(Q) \end{bmatrix}$$

The system $\Psi(\mathbf{V}) = \Pi$ is equivalent to:

$$\begin{aligned} P(\mathbf{V}_1) + P(\mathbf{V}_2) &= P(P) \\ P(\mathbf{V}_1) + P(\mathbf{V}_3) + P(\mathbf{V}_4) &= P(\neg P \vee Q) \\ P(\mathbf{V}_1) + P(\mathbf{V}_3) &= P(Q) \end{aligned}$$

Such that $\sum_{i=4}^{4} P(\mathbf{V}_i) = 1$. By doing some algebraic transformations, one obtains:

$$\begin{aligned} P(\neg P \vee Q) + P(P) - 1 - P(Q) &= P(\mathbf{V}_1) + P(\mathbf{V}_3) + P(\mathbf{V}_4) \\ &\quad + P(\mathbf{V}_1) + P(\mathbf{V}_2) - 1 - P(\mathbf{V}_1) - P(\mathbf{V}_3) \\ &= -P(\mathbf{V}_3) \leqslant 0 \end{aligned}$$

Then, one can write:

$$P(\neg P \vee Q) + P(P) - 1 \leqslant P(Q)$$

One also notices that:

$$P(Q) - P(\neg P \vee Q) = P(\mathbf{V}_1) + P(\mathbf{V}_3) - P(\mathbf{V}_1) - P(\mathbf{V}_3) - P(\mathbf{V}_4) = -P(\mathbf{V}_4) \leqslant 0$$

In conclusion, we obtain:

$$\max\{0, P(\neg P \vee Q) + P(P) - 1\} \leqslant P(Q) \leqslant P(\neg P \vee Q).$$

Remark 28.1
The bounds relative to $P(Q)$ are not found by a geometric approach, but by an algebraic approach.

The geometric interpretation is based on the following theorem.

Theorem 28.1
Let X and Y be two normed spaces and Ψ a linear mapping from X onto Y. If $\{\mathbf{V}_i\}$ is a set of extreme vectors according to the norm defined in X and Y, then the set $\{\Psi(\mathbf{V}_i)\}$ is also a set of extreme vectors according to the same norm.

Proof
Ψ is a linear operator. ■

THE USE OF THE INFINITE NORM: $L_\infty(x) = \max |x_i|$

EXTREME VECTORS

We express in the space X (normed space of possible worlds) the extreme vector concept by the fact that the norm of any vector $\mathbf{V}$ is equal to 1, which can be written

as $\|\mathbf{V}\|_\infty = 1$. Thus, we can write $\max|\mathbf{V}_i| = 1$ which is equivalent to $\max|P(\mathbf{V}_i)| = 1$, $\forall i \in \{1, 2, 3, 4\}$. However, since $\sum P(\mathbf{V}_i) = 1$, we know that if one component of the vector $\mathbf{V}$ is equal to 1 then the others must be 0. Finally, we have obtained four extreme vectors (according to $\|\cdot\|_\infty$) corresponding to the standard basis for $\mathbb{R}^4$ (the classical Euclidean space):

$$\mathbf{V}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{V}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{V}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{V}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

By using the linear relation due to the product of the consistent matrix $\boldsymbol{C}$ by the column vector of possible worlds probabilities $\mathbf{V}$, we can write $\boldsymbol{\Psi}(\mathbf{V}) = \boldsymbol{\Pi}$, which is equivalent to $\boldsymbol{C}\mathbf{V} = \boldsymbol{\Pi}$. Using Theorem 28.1, $\boldsymbol{\Psi}(\mathbf{V})$ is also extreme in the space of sentences according to the same norm.

This algebraic property enables us to plot the consistent region according to the infinite norm. As noted above, the extreme vectors in the space of possible worlds correspond to the standard basis for $\mathbb{R}^4$, so the extreme vectors in the space of sentences are the column vectors of the matrix $\boldsymbol{C}$ (consistent):

$$\boldsymbol{\Pi}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \boldsymbol{\Pi}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \boldsymbol{\Pi}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{\Pi}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Geometrically, the consistent region is the inside of the convex hull of the extreme vectors $\psi(\mathbf{V})$. Since these extreme vectors are the columns of the matrix $\boldsymbol{C}$, we can plot the consistent region using $\|\cdot\|_\infty$ as in Fig. 28.2.

The points outside this volume are inconsistent. When using the mixed product in order to determine the plane equations, the constraints corresponding to the consistent

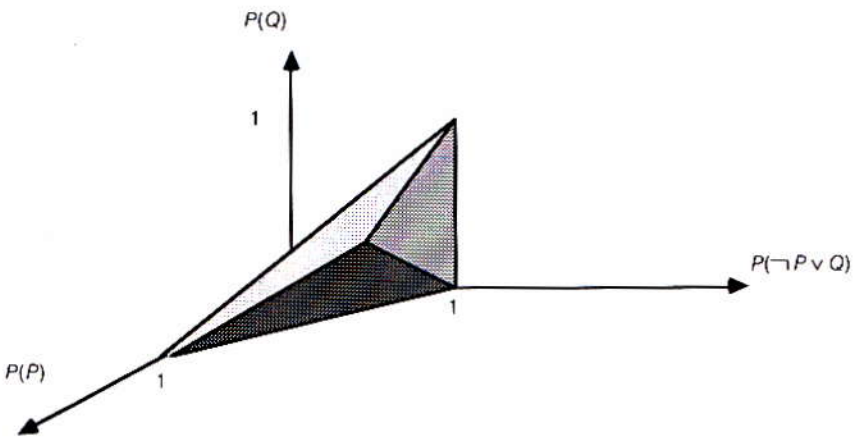

Figure 28.2 The interior of this convex hull is the consistent region.

region are easily found. Thus, we obtain the inequalities:

$$P(Q) \leqslant P(\neg P \vee Q)$$
$$P(\neg P \vee Q) + P(P) - 1 \leqslant P(Q)$$

The length of the interval corresponding to $P(Q)$ is equal to: $I_\infty = 1 - P(P)$. It depends only on the value of $P(P)$.

Remark 28.2
The bounds relative to $P(Q)$ are now found geometrically. The geometric region has been searched in order to match with the region found algebraically. The geometrical construction according to the infinite norm is a particular semantic.

THE CONSISTENT REGION DEDUCED FROM ISOMETRIES

It is of interest to see what is the consistent region when dealing with special automorphisms, called *isometries*.

Definition 28.1
Let $f: X \to Y$ a non-singular mapping and X and Y be two normed spaces. f is called an isometry if $\|x - y\| = \|f(x) - f(y)\| \; \forall (x, y) \in X^* Y$.

Proposition 28.1
If $\boldsymbol{C} \in L(\mathbb{R}^n)$, the following assertions are equivalent:

(a) $\boldsymbol{C}$ is an isometry
(b) $\|\boldsymbol{C}^*\mathbf{V}\|_\infty = \|\mathbf{V}\|_\infty \; \forall \mathbf{V} \in \mathbb{R}^n$

Proof
We apply Definition 28.1 with $V = (x - y)$ and f associated with $\boldsymbol{C}$. ■

Remark 28.3
Any extreme vector $\boldsymbol{\Pi} = \boldsymbol{C}^*\mathbf{V}$ defined in the space of sentences is an image of an extreme vector $\mathbf{V}$ defined in the space of possible worlds. However, if $\boldsymbol{C}$ is an isometry, the second equation of Proposition 28.1 can be written as: $\|\boldsymbol{\Pi}\|_\infty = 1$, where $\boldsymbol{\Pi}$ is an extreme vector. Thus, the method used in order to find $P(Q)$, when $P(P)$ and $P(\neg P \vee Q)$ are known, is less complicated. This is due to the fact that any solution for $\mathbf{V}$ has to check equation (b) of Proposition 28.1.

THE USE OF THE L_1 NORM: $L_1(x) = \sum |x_i|$

Using the property of a linear mapping and the continuity concept, we define an overvalue relative to the norm of any vector in the space of possible worlds. The consistent region is contained inside the maximal area assigned to this overvalue.

THE MAXIMAL AREA CONTAINING THE CONSISTENT REGION

In the case of the L_1 norm, an extreme vector in the space of possible worlds is defined by $\|\mathbf{V}\|_1 = 1$. Therefore, by this definition, all the vectors defined in the space of possible worlds become extreme. This remark is inherent to the fact that a probability must satisfy the concept of exclusivity and exhaustivity.

Additional information is required in order to obtain a square matrix $\boldsymbol{C}$. Therefore, the problem is approached via Proposition 28.2.

Proposition 28.2
The image of a compact set by a continuous mapping is also a compact set in the space of sentences.

Proof
See Choquet (1964). ■

As $\mathbb{R}^4$ has finite dimension, the compact image is in particular bounded, we can express the continuity of the matrix $\boldsymbol{C}$ by:

$$\|\boldsymbol{C}\mathbf{V}\|_1 \leqslant \|\boldsymbol{C}\|_1 \|\mathbf{V}\|_1 \forall \mathbf{V} \in \mathbb{R}^4$$

The norm of a linear mapping in a normed space is defined by the following relation:

$$\|\boldsymbol{C}\|_1 = \max \|\boldsymbol{C}\mathbf{V}\| 1 / \|\mathbf{V}\|_1$$

and the extreme condition is expressed by $\|\mathbf{V}\|_1 = 1$. Thus, we obtain:

$$\|\boldsymbol{C}\|_1 = \max \|P(P); P(\neg P \vee Q); P(Q); P(\tau)\|_1 = 4$$

In conclusion, the final result is:

$$\|\boldsymbol{C}\mathbf{V}\|_1 \leqslant \|\boldsymbol{C}\|_1 = 4$$

which is equivalent to:

$$P(P) + P(\neg P \vee Q) + P(Q) + P(\tau) \leqslant 4$$

(4 is the number of sentences). The maximal area is then:

$$\mathscr{D} = \{[P(P), P(\neg P \vee Q), P(Q), P(\tau)]^{\mathrm{T}} \text{ such that } P(P) + P(\neg P \vee Q) + P(Q) + P(\tau) \leqslant 4\}$$

where T indicates a vector transpose. The consistent region is strictly contained inside the region $\mathscr{D}$.

THE CONSISTENT REGION

Proposition 28.3
Any vector defined in the space of possible worlds is extreme according to this norm.

Proof
This is due to the fact that any probability vector must satisfy $\sum P(\mathbf{V}_i) = 1$. ■

Before giving the method and a solution, we present some topological properties.

$$\text{Let } \mathscr{P} = \{\mathbf{V} = [P(\mathbf{V}_1), P(\mathbf{V}_2), \ldots, P(\mathbf{V}_n)]^{\mathrm{T}} / \sum P(\mathbf{V}_i) = 1 \text{ and } \forall i,\ P(\mathbf{V}_i) = P_i \geqslant 0\}$$

Lemma 28.1
$\mathscr{P}$ is the space of possible worlds and it is a convex and compact set in $\mathbb{R}^n$.

Proof
$\mathscr{P}$ is convex and closed because it is the intersection of closed convex sets; it is also bounded because it is contained inside the unit sphere $\|\mathbf{V}\|_1 = 1$. ■

Theorem 28.2
Let U be a linear operator in $\mathbb{R}^n$, $\boldsymbol{C} = (C_{ij})$ its associated matrix relative to the standard basis for $\mathbb{R}^n$. If $\sum C_{ij} = n\,\forall j$ ($n-1$ is the number of sentences in the set of beliefs), $C_{ij} \geqslant 0$, then we can write the following assertions:

(a) $U(\mathscr{P}) \subset n\mathscr{P}$
(b) $\|\boldsymbol{\Pi}\|_1 = n\|\mathbf{V}\|_1$. $\simeq \mathcal{M}$

Proof
By using the hypothesis of this theorem, one can write:

$$\|\boldsymbol{C}\mathbf{V}\|_1 = \sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}\mathbf{V}_j = \sum_{j=1}^{n}\sum_{i=1}^{n} C_{ij}\mathbf{V}_j = \sum_{j=1}^{n}\mathbf{V}_j \sum_{i=1}^{n} C_{ij} = n\sum_{j=1}^{n}\mathbf{V}_j$$

and both of the relations (a) and (b) are proved. ■

Proposition 28.4
The sets $\mathscr{P}$, $U(\mathscr{P})/n$, $U^2(\mathscr{P})/n^2, \ldots, U^k(\mathscr{P})/n^k, \ldots$, form a decreasing sequence with an intersection D different from $\varnothing$.

Proof
$U^k(\mathscr{P})/n^k$ is a decreasing sequence when applying Theorem 28.2, so any finite intersection of the squence is different from $\varnothing$. Now, for all k, U^k is continuous because it is a linear operator in a finite dimension space, thus, $U^k(\mathscr{P})/n^k$ is compact, and $\cap U^k(\mathscr{P})/n^k = D \neq \varnothing$. ■

Proposition 28.5
$U(D) = D$ and for each probability vector $\mathbf{V}$, $U(\mathbf{V}) \in D$.

Proof
If $\mathbf{V} \in D$, for all $k \geqslant 0$, $\mathbf{V} \in U^k(\mathscr{P})/n^k$, therefore $U(\mathbf{V}) \in U^{k+1}(\mathscr{P})/n^k$, e.g. $U(\mathbf{V}) \in D$. If $\mathbf{W} \in D$, by the definition of D, for all $k \geqslant 0$, $\exists \mathbf{V}_k \in U^k(\mathscr{P})/n^k$ such that $\mathbf{W} = U(\mathbf{V}_k)$. As $\mathscr{P}$ is compact, the sequence $\mathbf{V}_k$ has at least one adherence value $\mathbf{V}$, then $\mathbf{V} \in D$, since $V_k \in \cap U^i(\mathscr{P})/n^i$ with $i \leqslant k$ and the $U^i(\mathscr{P})/n^i$ are closed and decreasing. In conclusion, as $U(D) \subset D$ and $D \subset U(D)$ then $U(D) = D$. ■

Theorem 28.3
If $\boldsymbol{C}(k) = (C_{ij}(k))$ is the matrix associated with U^k/n^k relative to the canonical basis is for $\mathbb{R}^n$, then a strictly increasing sequence, $k_1, k_2, k_3, \ldots$ of natural numbers exists so that, for all i and j, the sequence $(C_{ij}(k_p))$ tends to the limit matrix $\boldsymbol{D} = (D_{ij})$ as p tends to infinity and $L(\mathscr{P}) = \boldsymbol{D}$, where L is a linear operator associated to the matrix $\boldsymbol{D} = (D_{ij})$.

Proof
For all i, j, k, we can write: $0 \leqslant C_{ij}^k \leqslant 1/n^k$; so, for all i, j, the sequence (C_{ij}^k) possesses an adherence value (Bolzano–Wiertrass). So by setting in order the finite set of indices i and J, we can, by n^2 extractions of the sequence, find an increasing sequence of natural numbers (k_p) so that, for all i and j, the sequence $(C_{ij}(K_p))$ tends to $\boldsymbol{D} = (D_{ij})$(adherence value) as p tends to infinity. For all $p \geqslant 1$, we can write, using the definition of L: $L(\mathscr{P}) \subset U^{k_p}(\mathscr{P})/n^{k_p}$, and, therefore, $L(\mathscr{P}) \subset \boldsymbol{D}$ because $U^i(\mathscr{P})/n^i$ is a decreasing sequence. Let $\mathbf{V} \in \boldsymbol{D}$; for all natural numbers p, we can find $\mathbf{V}_{k_p} \in \mathscr{P}$ such that $\mathbf{V} = U^{k_p}(\mathbf{V}_{k_p})/n^{k_p}$. If $\mathbf{W}$ is an adherence value associated to the sequence $\mathbf{V}_{kp}$, we obtain the result by using, for example, the following inequality:

$$\| U^{k_p}(\mathbf{V}_{k_p})/n^{k_p} - L(\mathbf{W}) \| \leqslant \| U^{k_p}(\mathbf{V}_{k_p})/n^{k_p} - L(\mathbf{V}_{k_p}) \| + \| L(\mathbf{V}_{k_p}) - L(\mathbf{W}) \| \qquad \blacksquare$$

Proposition 28.6
$\mathscr{P}$ is the convex hull relative to $\{\mathbf{e}_i\}$, D is the convex hull relative to $\{L(\mathbf{e}_i)\}$, where $\{\mathbf{e}_i\}$ is the canonical basis associated with $\mathbb{R}^n$.

Proof
This is due to the fact that L is a linear operator and $L(\mathscr{P}) = \boldsymbol{D}$. $\blacksquare$

APPLICATION OF THE PREVIOUS TOPOLOGICAL RESULTS

The original problem was $\boldsymbol{C}\mathbf{V} = \boldsymbol{\Pi}$, where $\boldsymbol{C}$ is the consistent matrix, $\mathbf{V}$ is the probability vector in the space of possible worlds, and $\boldsymbol{\Pi}$ is the vector probability associated with the sentences. Methods using the entropy of the distribution $\mathbf{V}$ have been used, but they remain approximate and subjective. Our consistent matrix is:

$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

We complete the matrix $\boldsymbol{C}$ by adding another row so that the sum of each column will be equal to the number of sentences $+ 1 = 4 = n$, in this example. Our new matrix, $\boldsymbol{C}'$, is square and all its elements values are positive. It corresponds to our linear operator U used in Theorem 28.3, and is as follows:

$$\boldsymbol{C}' = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ [1 & 3 & 2 & 3] \end{bmatrix}$$

The problem becomes:

$$\boldsymbol{C}'^{*}\mathbf{V} = \boldsymbol{\Pi}' = [P(P), P(\neg P \vee Q), P(Q), \mathscr{A}]^{\mathrm{T}}$$

where $\mathscr{A}$ is the component deduced from the added row of the matrix $\boldsymbol{C}'$. In order to find $\boldsymbol{D}$ we have to multiply the square new matrix $\boldsymbol{C}'^{*}\ 1/n$ by itself p times; the number p of iterations used is the **precision** of the method.

We obtain the following diagram:

$$\boldsymbol{C} \rightarrow \boldsymbol{C}'^{*}\ 1/n \rightarrow \boldsymbol{C}'^{2*}1/n^2 \rightarrow \boldsymbol{C}'^{3*}1/n^3 \rightarrow \cdots \rightarrow \boldsymbol{C}'^{p*}1/n^p$$

$\boldsymbol{C}'^{p*}1/n^p$ corresponds to the limit matrix $\boldsymbol{D} = (D_{ij})$ of the topological results; this matrix is associated with the linear operator L. As the approximated limit matrix $\boldsymbol{C}'^{p*}1/n^p$ is found, then $\boldsymbol{D}$ is obtained by the convex hull $[\boldsymbol{C}'^{p*}1/n^p](\mathbf{e}_i)$, where $\{\mathbf{e}_i\}$ is the canonical base of our vector space. As $[\boldsymbol{C}'^{*}1/n](\boldsymbol{D}) = \boldsymbol{D}$, inside $\boldsymbol{D}$, which is completely defined, any point considered is consistent. However, to find $P(Q)$, we have to project to its corresponding axis. The programming of this method is easy to define, as any iterative process.

THE CONSISTENT REGION DEDUCED FROM ISOMETRIES

As each vector defined in the space of possible worlds is extreme according to this norm, when dealing with isometries, we obtain the following proposition.

Proposition 28.7
If $\boldsymbol{C} \in L(\mathbb{R}^n)$, the following assertions are equivalent:

(a) $\|\boldsymbol{C}^{*}\mathbf{V}\|_1 = \|\boldsymbol{\Pi}\|_1 = \|\mathbf{V}\|_1 = 1$
(b) $\sum C_{ij} = 1\ \forall j \in \{1, \ldots, n\}$ and $C_{ij} \geqslant 0$.

Proof
This is Theorem 28.2 with $n = 1$. ■

Remark 28.4
This is a special case ($n = 1$) of the work done previously. However, any vector $\boldsymbol{\Pi} = (\boldsymbol{\Pi}_i)$ satisfies $\sum \boldsymbol{\Pi}_i = 1$. This case is the one where the space of sentences is a probability space. The solution for $P(Q)$ is unique and is equal to $1 - P(\neg P \vee Q) - P(P) - \mathscr{A}$.

Remark 28.5
The approximate methods using the entropy concept (Nilsson, 1986) in order to find a solution for $P(Q)$ can now be checked. We can apply them to isometries and the result for $P(Q)$ must be $1 - P(\neg P \vee Q) - P(P) - \mathscr{A}$. However, some work needs to be done in order to define all 'logical isometries'; one certainly has to change the binary representation of the truth values assigned to predicates.

THE USE OF THE L_2 NORM: $L_2(x) = (\sum x_i^2)^{1/2}$

In this section, we use the Euclidean norm to find the minimal sphere containing the 'consistent' region. We maximize a function which is defined as the image of an extreme vector in the space of possible worlds. We show that the problem is equivalent to finding the eigenvalues associated with a quadratic form.

THE MINIMAL SPHERE CONTAINING THE CONSISTENT REGION

This approach is based on weights W_i (Zadeh, 1975; 1983) given to the sentences instead of probabilities and the concept of continuity in the space of finite dimensions. Each weight W_i is assigned to a possible world. One condition is that each weight belongs to the interval [0, 1] and their sum can be greater than 1. By using this norm, we are approaching a fuzzy logic.

To make calculations easier, in order to apply some methods using spectral theory, one can add additional information τ (tautology) to fulfil the necessary hypothesis. As $\|\mathbf{V}\|_2^2$ cannot be greater than 1, we impose the condition that the extreme vector is defined according to this norm as:

$$\|\mathbf{V}\|_2^2 = 1, \text{ which is equivalent to } \sum W_i^2 = 1 \tag{28.1}$$

We are interested in determining the minimal sphere containing the consistent region. The constraint corresponding to this aim is: $\exists k_0 \in \mathbb{R}^+$ such that:

$$\|\boldsymbol{\Pi}\|_2^2 = k_0 \tag{28.2}$$

The continuity of the homomorphism corresponding to the consistent matrix yields the following assertion:

$$\exists K > 0;\ \|\boldsymbol{C}^*\mathbf{V}\|_2 \leqslant K\|\mathbf{V}\|_2 \Leftrightarrow \|\boldsymbol{\Pi}\|_2 \leqslant K \quad (\text{as } \|\mathbf{V}\|_2 = 1) \tag{28.3}$$

Equation (28.2) is equivalent to the sphere whose equation in $\mathbb{R}^4$ is:

$$W^2(P) + W^2(\neg P \vee Q) + W^2(Q) + W^2(\tau) = k_0$$

We choose k_0 equal to $\inf\{K_i\}$ where K_i is the set of positive numbers satisfying the relation (28.3), and τ is a tautology added as the last row vector in the square consistent matrix. Finally, as $\boldsymbol{C}^*\mathbf{V} = \boldsymbol{\Pi}$, our optimization approach is to solve according to $W(\mathbf{V}_i) = w_i$ the following problem:

$$\max f(w_1, w_2, w_3, w_4) = \max\{(w_3 + w_4)^2 + (w_1 + w_2 + w_4)^2 + (w_2 + w_4)^2 + (w_1 + w_2 + w_3 + w_4)^2\}$$

subject to:

$$Q(w_1, w_2, w_3, w_4) = \sum w_i^2 = 1 \qquad i \in \{1, 2, 3, 4\}$$

The mapping Q is quadratic and positive definite, so solutions can be found by using spectral theory properties. Before giving a solution using spectral properties, one can transform our optimized problem into an equivalent one. In order to obtain a well-known

function to optimize, we shall change the variables in the following manner:

$$(w_3 + w_4) = x_1;\ (w_1 + w_2 + w_4) = x_2;\ (w_2 + w_4) = x_3;\ (w_1 + w_2 + w_3 + w_4) = x_4$$

The determinant of the associated matrix $\mathscr{M}$ of this linear system is:

$$|\mathscr{M}| = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

This determinant is non-zero (1 in this example), thus the solution for w_i is unique. This transformation is quite independent of the number of sentences; it can be used each time the determinant of the matrix is different from 0 (the matrix is invertible).

The problem is transformed into the following one:

$$\max f(x_1, x_2, x_3, x_4) = \max \sum x_i^2$$

subject to:

$$Q(x_1, x_2, x_3, x_4) = (x_2 - x_3)^2 + (x_4 + x_3 - x_1 - x_2)^2 + (x_4 - x_2)^2 + (x_1 + x_2 - x_4)^2 = 1$$

Let us consider $\mathbb{R}^4$ with its Euclidean structure. We associate with the quadratic form Q, for example, the symmetric automorphism U defined by: $Q(x) = (U(x), x) = (x, U(x))$, the comma denoting the scalar product. This automorphism may make the problem less complicated to solve. Therefore, the automorphism which can be associated with the quadratic form Q is not unique.

The differential of

$$\begin{aligned} g\colon\ \mathbb{R}^4 &\to \mathbb{R} \\ x &\to Q(x) - 1 \end{aligned}$$

is the following mapping:

$$h \to Q'(x) \cdot h = 2(U(x), h)$$

The differential of

$$\begin{aligned} f\colon\ \mathbb{R}^4 &\to \mathbb{R} \\ x &\to (x, x) \end{aligned}$$

is the following mapping:

$$h \to f'(x) \cdot h = 2(x, h)$$

Theorem 28.4

Let $f\colon \boldsymbol{D} \to \mathbb{R}$, where $\boldsymbol{D}$ is an open set of $\mathbf{E}$ (vector space). If f possesses a local extremum for $x_0 \in \boldsymbol{D}$, then $f'(x_0) = 0$.

Conversely, if $f \in C^2$ in $\boldsymbol{D}$, and if $x_0 \in \boldsymbol{D}$ exists where $f'(x_0) = 0$ and $f''(x_0)$ is a positive definite (negative definite) quadratic form, then f possesses a local minimum (maximum) for x_0.

Proof
See Choquet (1964) and Dieudonné (1960). ■

Proposition 28.8
Let $f, g: \boldsymbol{D} \rightarrow \mathbb{R}$ differentiable in the open $\boldsymbol{D}$ of $\mathbf{E}$ (vector space), and let $V = \{x \in \boldsymbol{D} | g(x) = 0\}$ if $f | V$ possesses an extremum for x_0. Then the forms $f'(x_0)$ and $g'(x_0)$ are proportional.

Proof
See Dieudonné (1960). ■

Proposition 28.9
If $E = \mathbb{R}^P$, this proportionality is expressed by the proportionality of partial derivatives.

Proof
Since $f'(x_0)$ and $g'(x_0)$ are proportional, then it is easy to obtain the proportionality of the partial derivatives. ■

Proposition 28.8 proves that there exists $\lambda \in \mathbb{R}$ such that $(U(x), h) = \lambda(x, h)$ for all h. Thus, by Proposition 28.9, we deduce:

$$\exists \lambda \in \mathbb{R} \text{ with: } U(x) = \lambda x$$

The eigenvalues are all positive (Q is positive definite), the extreme vectors are obtained in all points x which are eigenvectors of the automorphism U with $Q(x) = 1$. In those solution points for example x_0, one can write:

$$f(x_0 + h) = (x_0, x_0) + 2(x_0, h) + (h, h)$$
$$Q(x_0 + h) = Q(x_0) + 2(U(x_0), h) + Q(h)$$

with $Q(x_0) = Q(x_0 + h) = 1$. Thus, we obtain the following relations:

$$(x_0, h) = (U(x_0), h)^* 1/\lambda = -Q(h)/2\lambda$$

and thus we can write:

$$f(x_0 + h) = f(x_0) + (h, h) - Q(h)/\lambda$$

Since λ_m and λ_M are respectively the minimal and maximal eigenvalue, we have, for all h,

$$\lambda_m(h, h) \leqslant Q(h) \leqslant \lambda_M(h, h)$$

for $\lambda = \lambda_m$ we obtain, for all h,

$$f(x_0 + h) \leqslant f(x_0)$$

and thus the function f possesses a maximum, and this minimal eigenvalue is the solution to our problem. So $\lambda = \lambda_m = k_0 = \max f(x_1, x_2, x_3, x_4)$, and for $\lambda = \lambda_M$ we obtain a minimum. At any point obtained $(x_0, x_0) = (U(x_0)/\lambda, x_0) = (1/\lambda)Q(x_0) = 1/\lambda$.

We should, however, select among all possible solutions those which give a weight value in the interval [0, 1]; so we must have the following inequalities:

$$\begin{cases} 0 \leqslant x_2 - x_3 \leqslant 1 \\ 0 \leqslant x_4 + x_3 - x_1 - x_2 \leqslant 1 \\ 0 \leqslant x_4 - x_2 \leqslant 1 \\ 0 \leqslant x_1 + x_2 - x_4 \leqslant 1 \\ 0 \leqslant x_i \leqslant 1 \end{cases}$$

If the eigenvector corresponding to the minimal eigenvalue does not agree with the previous inequalities, we shall have to test the eigenvector associated to the next minimal eigenvalue. The solution of our problem is transformed to the search of the eigenvalues of the following symmetric automorphism U deduced from the previous expression of Q (Gaussian method):

$$U = \begin{bmatrix} 2 & 2 & -1 & -2 \\ 2 & 4 & -2 & -3 \\ -1 & -2 & 2 & 1 \\ -2 & -3 & 1 & 3 \end{bmatrix}$$

The characteristic polynomial of this automorphism is:

$$P(\lambda) = \lambda^4 - 11\lambda^3 + 21\lambda^2 - 11\lambda + 1$$

Many numerical algorithms exist (Ciarlet, 1983) for determining the minimal eigenvalue. As has been proved, this minimum is equal to $k_0 \leqslant 1$. The consistent region is limited by the positive part of the sphere in which the radius is equal to $(\inf \lambda_i)^{1/2}$. It can be represented in $\mathbb{R}^3$ by Fig. 28.3.

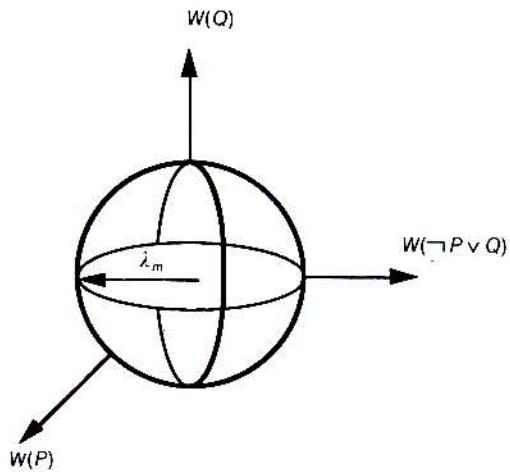

Figure 28.3 The interior of the positive part of this sphere is the minimal area containing the consistent region.

THE CONSISTENT REGION DEDUCED FROM ISOMETRIES

The idea has to do with the following question: in what conditions could the consistent region be contained inside the unit ball, which is the unit sphere in three dimensions? Mathematically this is expressed by the following equality:

$$\| C^* V \|_2^2 = \| V \|_2^2 = 1 \Leftrightarrow \| \Pi \|_2^2 = 1$$

Proposition 28.10
If $C \in L(\mathbb{R}^n)$ the following assertions are equivalent:

(a) C is an orthogonal matrix.
(b) $\| CV \|_2 = \| V \|_2$ (C is called an isometry).

Proof
$\| CV \|_2^2 = (CV, CV) = (V, C^* CV)$ as $C^* C = I$ (C is an orthonormal matrix) thus: $\| CV \|_2^2 = (V, V) = \| V \|^2$ and $\| CV \|_2 = \| V \|_2$. ■

Remark 28.6
Statement (a) of Proposition 28.10 is the characterization of the isometries. We know that in the case of $\mathbb{R}^2$, the rotations and the orthogonal symmetries are those matrices. In the case of $\mathbb{R}^4$, which is our example, not only are the family rotations obtained but other families as well.

Proposition 28.11
If $C \in L(\mathbb{R}^n)$ is an isometry then C^{T} remains an isometry. The set of all isometries is a group for the product operation.

Proof
$(C^* C)^{\mathrm{T}} = I = C^{\mathrm{T}} (C^*)^{\mathrm{T}} = C^{\mathrm{T}} (C^{\mathrm{T}})^* = I$; this proves that C^{T} is an orthogonal matrix which is equivalent to an isometry. In the general case, the following kind of matrices are isometries:

$$\begin{array}{c} \qquad\qquad\qquad\quad p \qquad\quad q \\ \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \ldots 0 \ldots 0 \\ 0 & 1 & 0 & 0 \ldots & 0 \ldots & 0 \ldots 0 \ldots 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 \ldots 0 & & & \cos\theta \ldots & \sin\theta \ldots & 0 \ldots 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 \ldots 0 & & & -\sin\theta \ldots & \cos\theta \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 \ldots 0 & & & 0 \ldots & & 1 \end{bmatrix} \end{array}$$

where p and q are subject to the condition:

$$1 \leqslant p < q \leqslant n$$

p and q are not only indications of columns but also indications of lines, n is the

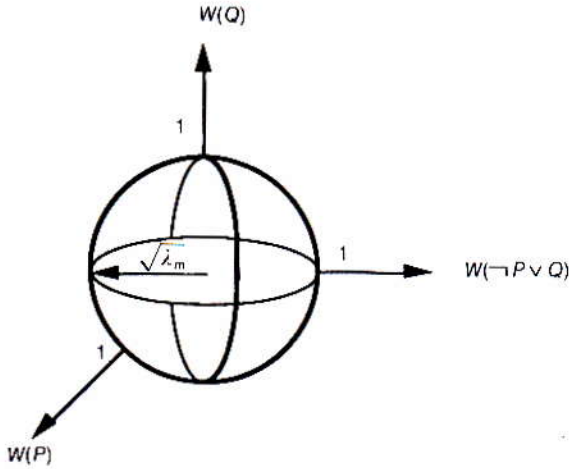

Figure 28.4 The interior of the positive part of this sphere is the minimal area containing the consistent region.

dimension of the matrix, θ is the angle of the vector of indices p and q of the rotation. The minimal area containing the 'consistent region' can be represented in $\mathbb{R}^3$ by Fig. 28.4.

A GEOMETRICAL INTERPRETATION OF BAYES'S THEOREM

Bayes inference (Duda *et al.*, 1976; Glymour, 1985) allows of easy geometrical interpretation when using the norm L_1. We show that the conditional probabilities $P(S_i/\boldsymbol{D})$ are given by the boundary of the unit ball associated to the L_1 norm.

Theorem 28.5

If the set $S_1, S_2, S_3, \ldots, S_n$ realizes a partition of the fundamental space $\boldsymbol{\Omega}$, e.g. for all $(i, j) S_i \cap S_j = \varnothing$ if $i \neq j$ and $\cup S_i = \boldsymbol{\Omega}$, and $D \subset \mathscr{P}(\boldsymbol{\Omega})$ is an event, then we can write, for all i, the following formula:

$$P(S_i/D) = \frac{P(D/S_i)P(S_i)}{\sum_{i=1}^{i=n} P(D/S_i)P(S_i)}$$

and the equality is true for all i in $[1, n]$.

We can express this in vector notation as:

$$\begin{bmatrix} P(S_1/D) \\ P(S_2/D) \\ P(S_3/D) \\ \vdots \\ P(S_n/D) \end{bmatrix} = \frac{1}{\sum_{i=1}^{i=n} P(D/S_i)P(S_i)} \begin{bmatrix} P(D/S_1)P(S_1) \\ P(D/S_2)P(S_2) \\ P(D/S_3)P(S_3) \\ \vdots \\ P(D/S_n)P(S_n) \end{bmatrix}$$

One notices that if we call $\boldsymbol{Y} = (\boldsymbol{Y})_i = P(S_i/D)$, the conditional vector associated with D, and $\boldsymbol{X} = (\boldsymbol{X})_i = P(D/S_i)P(S_i)$, this equation can be written as:

$$\boldsymbol{Y} = \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|_1}$$

and thus $\|\boldsymbol{Y}\|_1 = 1$. The vector $\boldsymbol{Y}$ belongs to the boundary of the unit ball associated to the L_1 norm. Other results could be deduced by this way of interpretation, the norm L_1 seems to be much more adequate to inference concept using probabilities.

CONCLUSION

In this chapter we have presented different ways of obtaining the consistent region. We pointed out that the notion of consistency is inherent in the choice of a topology. One notices that the notion of inclusion between the set of possible worlds and the set of sentences is hidden when using the infinite norm. When dealing with the 'logical isometries' (logical consistent isometries), the solution for $P(Q)$ is easy to find. The use of the L_1 norm is much more advantageous in the geometrical aspect than that of the infinite norm, because many relations exist between the linear mappings and the L_1 norm. The consistent region when using this norm can be reduced, and this is not a mathematical advantage but an experimental advantage, since we must distinguish between the possible realizations of the proposition $P(Q)$. Experimentally, the ideal case is that corresponding to a unique solution for $P(Q)$. The condition of isometry when using the L_1 norm is equivalent to the exclusivity and exhaustivity of the sentences: it is the case where the space of sentences is a probability space. The determination of the probability associated with a sentence S given the other sentences is then unique. The case of the L_2 norm requires a definition of weights in the space of possible worlds. Our approach is based on the minimal area containing the consistent region corresponding to this norm. The results obtained are interesting since we reach the spectral theory which is still a very rich area of mathematics. Some isometries have been cited; their minimal area containing the consistent region is the unit sphere. The geometric interpretation of Bayes's theorem is easily given by the boundary of the unit ball when using the L_1 norm.

REFERENCES

Choquet, G. (1964) *Cours d'analyse*, Vol. 2, Paris: Masson.

Ciarlet, P.G. (1983) *Introduction à l'analyse numérique matricielle et à l'optimisation*, Paris: Masson.

Dieudonné, J. (1960) *Foundations of Modern Analysis*, New York: Academic Press.

Duda, R.O., Hart, P. and Nilsson, N.J. (1976) Subjective Bayesian methods for rule-based inference systems, in *Proceedings of the 1976 National Computer Conference, AFIPS*, **45**, 1075–1082. Reprinted in B.W. Webber and N.J. Nilsson (eds) (1981) *Readings in Artificial Intelligence*, Palo Alto, CA: Tioga.

Glymour, C. (1985) Independence assumptions and Bayesian updating. *Artificial Intelligence*, **25**, 95–99.

Grosof, B.N. (1985) An inequality paradigm for probabilistic knowledge. In *Proceedings*

AAAI/IEEE Workshop on Uncertainty and Probability in Artificial Intelligence, Los Angeles.

Kleene, S.C. (1987) *Logique mathématique*, Paris: Jacques Gabay.

Nilsson, N.J. (1986) Probabilistic logic. *Artificial Intelligence*, **28**, 71–87.

Suppes, P. (1966) Probabilistic inference and the concept of two evidences, in J. Hintikka and P. Suppes (eds), *Aspects of Inductive Logic*, Amsterdam: North-Holland, pp. 49–65.

Zadeh, L.A. (1975) Fuzzy logic and approximate reasoning. *Synthese*, **30**, 407–428.

Zadeh, L.A. (1983) The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, **11**, 199–227.