Chapter 2

Belief as Probability

In this chapter we shall prove some simple properties of probability functions and consider some consequences of the assumption that Bel satisfies $(P1\text{–}2)$ for the question Q.

Throughout the chapter we assume that $Bel: SL \to [0,1]$ satisfies$(P1\text{–}2)$, that is, for $\theta, \phi \in SL$,

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

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

Notice that if V is a (two-valued) valuation on L in the usual sense of the propositional calculus then V satisfies $(P1\text{–}2)$ and hence is a probability function. In this sense belief as probability extends the classical two-valued semantics.

Proposition 2.1 *For* $\theta, \phi, \theta_1, \theta_2, ..., \theta_n \in SL$

(a) $Bel(\neg\theta) = 1 - Bel(\theta)$.

(b) If $\models \theta$ then $Bel(\neg\theta) = 0$.

(c) If $\theta \models \phi$ then $Bel(\theta) \leq Bel(\phi)$ and if $\models (\theta \leftrightarrow \phi)$ then $Bel(\theta) = Bel(\phi)$.

(d) $Bel(\theta \vee \phi) = Bel(\theta) + Bel(\phi) - Bel(\theta \wedge \phi)$.

More generally,

$$Bel(\bigvee_{i=1}^{n} \theta_i) = \sum_{\emptyset \neq S \subseteq \{1,...,n\}} (-1)^{|S|-1} Bel(\bigwedge_{i \in S} \theta_i).$$

Proof (a) follows directly from$(P1\text{–}2)$ since $\models \theta \vee \neg\theta$ and $\models \neg(\theta \wedge \neg\theta)$.

(b) now follows from $(P1)$ and (a).

(c) If $\theta \models \phi$ then $\models \neg(\theta \wedge \neg\phi)$ so, by $(P2)$,

$$1 \geq Bel(\theta \vee \neg\phi) = Bel(\theta) + Bel(\neg\phi) = Bel(\theta) + 1 - Bel(\phi)$$

by (a), and (c) follows. The second part now follows since if $\models (\theta \leftrightarrow \phi)$ then $\theta \models \phi$ and $\phi \models \theta$.

(d) By (c), $Bel(\theta \vee \phi) = Bel((\theta \wedge \neg\phi) \vee \phi)$, since $\models (\theta \vee \phi) \leftrightarrow ((\theta \wedge \neg\phi) \vee \phi)$. By ($P2$), this equals $Bel(\theta \wedge \neg\phi) + Bel(\phi)$. Also, by (c) and (P2) respectively, $Bel(\theta) = Bel((\theta \wedge \neg\phi) \vee (\theta \wedge \phi)) = Bel(\theta \wedge \neg\phi) + Bel(\theta \wedge \phi)$. Removing $Bel(\theta \wedge \neg\phi)$ gives the result.

The generalization is proved by induction on $n \geq 2$. Assume the result for $n-1$. Then, using (c) freely and collecting summands which do/do not contain $(\theta_{n-1} \vee \theta_n)$ as a conjunct,

$$Bel\left(\bigvee_{i=1}^{n} \theta_i\right) = Bel\left(\bigvee_{i=1}^{n-2} \theta_i \vee (\theta_{n-1} \vee \theta_n)\right) =$$

$$\sum_{\emptyset \neq S \subseteq \{1,\ldots,n-2\}} (-1)^{|S|-1} Bel\left(\bigwedge_{i \in S} \theta_i\right) + \sum_{S \subseteq \{1,\ldots,n-2\}} (-1)^{|S|} Bel\left((\theta_{n-1} \vee \theta_n) \wedge \bigwedge_{i \in S} \theta_i\right).$$

Now since

$$\models ((\theta_{n-1} \vee \theta_n) \wedge \bigwedge_{i \in S} \theta_i) \leftrightarrow (\theta_{n-1} \wedge \bigwedge_{i \in S} \theta_i) \vee (\theta_n \wedge \bigwedge_{i \in S} \theta_i)$$

we can replace $Bel((\theta_{n-1} \vee \theta_n) \wedge \bigwedge_{i \in S} \theta_i)$ in this last expression by

$$Bel(\theta_{n-1} \wedge \bigwedge_{i \in S} \theta_i) + Bel(\theta_n \wedge \bigwedge_{i \in S} \theta_i) - Bel(\theta_{n-1} \wedge \theta_n \wedge \bigwedge_{i \in S} \theta_i)$$

and sorting out terms now gives the required right hand side in (d). □

Given a probability function Bel on SL it is completely standard to extend it (partially) to a conditional probability function as well by setting

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

whenever $Bel(\phi) \neq 0$, whilst leaving open this value when $Bel(\phi) = 0$. We shall adopt this convention henceforth. Notice that by (c) above, as $\theta \wedge \phi \models \phi$, $Bel(\theta \wedge \phi) \leq Bel(\phi)$ so for $Bel(\phi) \neq 0$, $0 \leq Bel(\theta \mid \phi) \leq 1$.

Furthermore since $Bel(\theta \wedge \phi) \leq Bel(\phi)$, if $Bel(\phi) = 0$ then $Bel(\theta \wedge \phi) = 0$ so that the requirement $Bel(\theta \mid \phi) = \dfrac{Bel(\theta \wedge \phi)}{Bel(\phi)}$ for $Bel(\phi) \neq 0$ is equivalent to $Bel(\phi) \cdot Bel(\theta \mid \phi) = Bel(\theta \wedge \phi)$ (whether or not $Bel(\phi) = 0$).

Proposition 2.2 *Fix $\phi \in SL$ such that $Bel(\phi) \neq 0$. Then, as a function of $\theta \in SL$, $Bel(\theta|\phi)$ is a probability function, $Bel(\phi|\phi) = 1$; if $Bel(\eta|\phi) \neq 0$ then $Bel(\eta \wedge \phi) \neq 0$ and for the corresponding conditional probability function for $Bel(\theta|\phi)$,*

$$\frac{Bel(\theta \wedge \eta|\phi)}{Bel(\eta|\phi)} = Bel(\theta|\eta \wedge \phi).$$

Proof If $\models \theta$ then $\models (\theta \wedge \phi) \leftrightarrow \phi$ so using (c) of 2.1

$$Bel(\theta|\phi) = \frac{Bel(\theta \wedge \phi)}{Bel(\phi)} = \frac{Bel(\phi)}{Bel(\phi)} = 1$$

giving $(P1)$. If $\models \neg(\theta \wedge \psi)$ then $\models \neg((\theta \wedge \phi) \wedge (\psi \wedge \phi))$ so

$$Bel(\theta \vee \psi|\phi) = \frac{Bel((\theta \vee \psi) \wedge \phi)}{Bel(\phi)} = \frac{Bel((\theta \wedge \phi) \vee (\psi \wedge \phi))}{Bel(\phi)}$$

$$= \frac{Bel(\theta \wedge \phi) + Bel(\psi \wedge \phi)}{Bel(\phi)} = Bel(\theta|\phi) + Bel(\psi|\phi)$$

The rest of the proposition follows easily by substituting in the definition. □

Notice that by the identity in proposition 2.2 and proposition 2.1(c) the result of successively conditioning on $\phi_1, \phi_2, \ldots, \phi_k$ is the same as conditioning, once, on $\bigwedge_{i=1}^{k} \phi_i$ and hence is independent of the order of $\phi_1, \phi_2, \ldots, \phi_k$.

Henceforth we shall make free use of propositions 2.1 and 2.2 and$(P1–2)$ without explicit mention.

The following theorem, due to Suppes [66], generalises proposition 2.1(c).

Theorem 2.3 *For $\phi, \theta_1, ..., \theta_n \in SL$, if $\theta_1, ..., \theta_n \models \phi$ then*

$$Bel(\phi) \geq 1 - \sum_{i=1}^{n} Bel(\neg\theta_i) = \sum_{i=1}^{n} Bel(\theta_i) - (n-1),$$

and furthermore this bound cannot be improved.

Proof The proof is by induction on n. For $n = 1$ the result is true by proposition 2.1(c). Assume the result for $n-1$. Then $\theta_1, ..., \theta_n \models \phi$ implies that $\theta_1, ..., \theta_{n-1} \models \neg\theta_n \vee \phi$ so, by inductive hypothesis,

$$Bel(\neg\theta_n \vee \phi) \geq 1 - \sum_{i=1}^{n-1} Bel(\neg\theta_i).$$

By proposition 2.1(d)

$$Bel(\neg\theta_n \vee \phi) = Bel(\phi) + Bel(\neg\theta_n) - Bel(\theta_n \wedge \phi) \leq Bel(\phi) + Bel(\neg\theta_n),$$

and from this the result follows. To see that, in general, this bound cannot be improved, take $\theta_1, \ldots, \theta_n, \phi$ to be tautologies. □

A Representation for Probability Functions

Taking L to be our 'default' language $\{p_1, p_2, ..., p_n\}$ let $\alpha_1, ..., \alpha_J$, where $J = 2^n$, run through the *atoms* of L, that is all conjunctions of the form

$$p_1^{\epsilon_1} \wedge p_2^{\epsilon_2} \wedge ... \wedge p_n^{\epsilon_n}$$

where $\epsilon_1, ..., \epsilon_n \in \{0,1\}$ and $p^1 = p,\ p^0 = \neg p$.

Then, by the disjunctive normal form theorem, for any $\phi \in SL$ there is a unique set

$$S_\phi \subseteq \{\alpha_1, ..., \alpha_J\}$$

such that

$$\models \phi \leftrightarrow \bigvee S_\phi$$

(where we take $\bigvee \emptyset = p_1 \wedge \neg p_1,\ \bigwedge \emptyset = p_1 \vee \neg p_1$ by convention).

It is easy to see that

$$S_\phi = \{\alpha_i |\ \alpha_i \models \phi\},$$

$$\models (\phi \leftrightarrow \theta) \ \Leftrightarrow\ S_\phi = S_\theta,$$

$$\models (\phi \rightarrow \theta) \ \Leftrightarrow\ S_\phi \subseteq S_\theta,$$

$$S_{\theta \wedge \phi} = S_\theta \cap S_\phi,$$

$$S_{\theta \vee \phi} = S_\theta \cup S_\phi,$$

$$S_{\neg\phi} = At^L - S_\phi,$$

where $At^L = \{\alpha_i |\ i = 1, ..., J\}$. Now since $\models \neg(\alpha_i \wedge \alpha_j)$ for $i \neq j$, repeated use of (P2) gives

$$Bel(\phi) = Bel(\bigvee S_\phi) = \sum_{\alpha_i \in S_\phi} Bel(\alpha_i) \quad \text{(even if } S = \emptyset).$$

Also, since $\models \bigvee_{i=1}^J \alpha_i$, $1 = \sum_{i=1}^J Bel(\alpha_i)$ and of course the $Bel(\alpha_i) \geq 0$.

From this it follows that Bel is completely determined by its values on the atoms α_i, i.e. by the vector

$$\langle Bel(\alpha_1), ..., Bel(\alpha_J)\rangle \ \in\ \mathbb{D}^L \text{ where } \mathbb{D}^L = \{\vec{x} \in \mathbb{R}^J |\ \vec{x} \geq 0,\ \sum_{i=1}^J x_i = 1\}.$$

Conversely given $\vec{a} \in \mathbb{D}^L$ define a function $Bel' : SL \rightarrow [0,1]$ by

$$Bel'(\phi) = \sum_{\alpha_i \in S_\phi} a_i.$$

Then for $\models \phi$, $S_\phi = \{\alpha_1, ..., \alpha_J\}$ and $Bel'(\phi) = \sum_{i=1}^J a_i = 1$. Also if $\models \neg(\theta \wedge \phi)$ then $S_{\theta\wedge\phi} = S_\theta \cap S_\phi = \emptyset$ so $S_{\theta\vee\phi}$ is the disjoint union of S_θ and S_ϕ and

$$Bel'(\theta \vee \phi) = \sum_{\alpha_i \in S_\theta \cup S_\phi} a_i = \sum_{\alpha_i \in S_\theta} a_i + \sum_{\alpha_i \in S_\phi} a_i = Bel'(\theta) + Bel'(\phi).$$

Finally $Bel'(\alpha_i) = a_i$ since $S_{\alpha_i} = \{\alpha_i\}$, so

$$\vec{a} = \langle Bel'(\alpha_1), ..., Bel'(\alpha_J)\rangle.$$

Summing up, what we have demonstrated here is that each probability function Bel (on SL) is determined by the point $\langle Bel(\alpha_1), ..., Bel(\alpha_J)\rangle \in \mathbb{D}^L$ and conversely every point $\vec{a} \in \mathbb{D}^L$ determines a unique probability function Bel satisfying

$$\langle Bel(\alpha_1), ..., Bel(\alpha_J)\rangle = \vec{a}.$$

This 1–1 correspondence between probability functions on L and points in $\mathbb{D}^L$ will be very useful in what follows. Indeed we shall frequently identify these notions, referring to a point in $\mathbb{D}^L$ as a probability function and conversely to a probability function as a point in $\mathbb{D}^L$.

Throughout this book α, α_i etc. will stand for atoms of L unless otherwise indicated.

A Reformulation of Question Q

Let K be as in the version of Q given at the end of Chapter 1, say K is the set of constraints

$$\sum_{j=1}^{r} c_{ji} Bel(\theta_j) = b_i \quad i = 1, ..., m$$

consistent with($P1$–2), i.e. having a solution satisfying also ($P1$–2).

Now replace each $Bel(\theta_j)$ in K by $\sum_{\alpha_i \in S_{\theta_j}} Bel(\alpha_i)$ and, in matrix notation, let

$$\langle Bel(\alpha_1), ..., Bel(\alpha_J)\rangle A_K = \vec{b}_K$$

be the resulting set of equations together with the equation

$$\sum_{i=1}^{J} Bel(\alpha_i) = 1.$$

Then if Bel is a probability function satisfying K, the point $\langle Bel(\alpha_1), ..., Bel(\alpha_J)\rangle$ from $\mathbb{D}^L$ satisfies

$$\vec{x} A_K = \vec{b}_K, \quad \vec{x} \geq 0.$$

Conversely if $\vec{a} \in \mathbb{R}^J$ satisfies these equalities and inequalities, then $\vec{a} \in \mathbb{D}^L$ and the corresponding probability function Bel such that $Bel(\alpha_i) = a_i$, $i = 1, ..., J$ satisfies K.

Summing up then, the natural correspondence between probability functions and points in $\mathbb{D}^L$ gives a correspondence between solutions of K satisfying $(P1\text{–}2)$ and points in

$$V^L(K) = \{\vec{x} \in \mathbb{R}^J \mid \vec{x}A_K = \vec{b}_K,\ \vec{x} \geq 0\} \subseteq \mathbb{D}^L.$$

The problem posed by Q, namely what value should we give to $Bel(\theta)$, for Bel satisfying K and $(P1\text{–}2)$, now becomes, what value should we give to $\sum_{\alpha_i \in S_\theta} x_i$, given that $\vec{x}A_K = \vec{b}_K$, $\vec{x} \geq 0$. Of course in this reformulated version the connection with uncertain reasoning seems to have been 'abstracted away'. Nevertheless it will often be useful to consider the reformulated version in what follows.

Notice that the matrix A_K, as presented here, has size $J \times (m+1)$ where m is the original number of equations in K. In practice it is almost always the case that m is much less than J $(= 2^n)$ so the system

$$\vec{x}A_K = \vec{b}_K,\quad \vec{x} \geq 0$$

rarely permits solving uniquely for $\vec{x}$ or indeed for $\sum_{\alpha_i \in S_\theta} x_i$, so that further assumptions beyond just$(P1\text{–}2)$ are needed in order to answer question Q. We shall consider some such assumptions in Chapter 6. For the present however we give a simple example, using the above representation.

Example
The quintessential example of our question Q is when

$$K = \{Bel(q) = b,\ Bel(p|q) = a\},\quad b > 0,\quad L = \{p, q\},$$

and we wish to give a value to $Bel(p)$. We consider this problem, and also the problem of assigning a value to $Bel(\neg p)$, under the assumptions $(P1\text{–}2)$.

In this case $Bel(p|q) = a$ reduces to $Bel(p \wedge q) = aBel(q) = ab$ (using the other constraint). The atoms of L are $p \wedge q$, $p \wedge \neg q$, $\neg p \wedge q$, $\neg p \wedge \neg q$ $(= \alpha_1, \alpha_2, \alpha_3, \alpha_4$ – in that order, say) and expanding the constraints in terms of the $Bel(\alpha_i)$ gives

$$Bel(p \wedge q) = ab,$$

$$Bel(q) = Bel((p \wedge q) \vee (\neg p \wedge q)) = Bel(p \wedge q) + Bel(\neg p \wedge q) = b,$$

whilst also

$$Bel(p \wedge q) + Bel(p \wedge \neg q) + Bel(\neg p \wedge q) + Bel(\neg p \wedge \neg q) = 1.$$

Hence setting $x_i = Bel(\alpha_i)$ there is a 1–1 correspondence between probability functions (on SL) satisfying K and $\langle x_1, x_2, x_3, x_4 \rangle$ satisfying

$$\langle x_1, x_2, x_3, x_4 \rangle \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \langle 1, ab, b \rangle, \qquad x_1, x_2, x_3, x_4 \geq 0.$$

The minimum value of $x_1 + x_2$ over these $\langle x_1, x_2, x_3, x_4 \rangle$ (equivalently the minimum possible value of $Bel(p \wedge q) + Bel(p \wedge \neg q) = Bel(p)$ for Bel a probability function satisfying K) is ab (when $x_1 = ab$, $x_2 = 0$, $x_3 = b(1-a)$, $x_4 = 1-b$). Arguing in this way we see that for a probability function Bel satisfying K, the range of consistent values for $Bel(p)$ is

$$1 - b + ab \geq Bel(p) \geq ab,$$

and for $Bel(\neg p)$

$$1 - ab \geq Bel(\neg p) \geq b(1-a).$$