Capitolo 5

LE ENTROPIE ETERODOSSE NELLA GESTIONE DELLA CONOSCENZA PARZIALE

1. Probabilità, possibilità, necessità

Nella terminologia probabilistica abituale i K elementi dell'alfabeto, o dello *spazio campionario*, $\mathcal{X} = \{x_1, x_2, ..., x_K\}$ si chiamano *eventi elementari* mentre i sottinsiemi di $\mathcal{X}$ si chiamano *eventi*.

L'identificazione "evento = sottinsieme dell'universo del discorso" risale a *The Doctrine of Chances*, del 1718, di Abraham de Moivre. Anche in questo capitolo ci limiteremo al caso finito: nella gestione dell'incertezza, ortodossa o eterodossa che sia, è sul finito che ci sono le idee pregnanti; in ogni caso per le applicazioni ai sistemi esperti basta.
Diremo *evento vuoto*, *evento pieno* ed eviteremo apposta i termini correnti, evento impossibile, evento certo, che creerebbero ambiguità.
Le probabilità ortodosse hanno (grossomodo) due interpretazioni, quella oggettiva, statistica, (limiti ideali di frequenze empiriche) e quella soggettiva, neo-bayesiana (gradi di fiducia misurati mediante la propensione a scommettere del soggetto: $Prob\{$il cambio euro-dollaro scende sotto il mezzo euro entro Natale$\} = \frac{2}{100} = 1 - \frac{98}{100}$ significa che sulla caduta dell'euro si è disposti a scommettere due contro novantotto, come all'ippodromo); è soprattutto questa seconda interpretazione che è interessante nelle applicazioni ai sistemi esperti.

Di solito le probabilità *bayesiane* (ortodosse, "normali") si assegnano sull'insieme finito $\mathcal{X}$ tramite un vettore di probabilità P a K componenti non negative:

$$P = (p_1, p_2, ..., p_K) \text{ dove } \sum_i p_i = 1$$
$$A \subseteq \mathcal{X}, \quad P(A) = \sum_A p_i$$

Alternativamente, le distribuzioni di probabilità sono individuate, sempre sul finito, dagli *assiomi*:

$$P(\emptyset) = 0, \quad P(A) \geq 0, \quad P(\mathcal{X}) = 1$$
$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

Esercizio. L'assioma $P(\emptyset) = 0$ è sovrabbondante (poni $B = \emptyset$ nell'ultimo assioma).

Indebolendo gli assiomi si ottengono le *distribuzioni di fiducia* (*confidence*, o *di consenso*):

$$\Phi(\emptyset) = 0, \quad \Phi(\mathcal{X}) = 1$$
$$A \subseteq B \Rightarrow \Phi(A) \leq \Phi(B)$$

(In questo caso l'assioma $\Phi(A) \geq 0$ più che sovrabbondante è proprio inutile). L'ultima riga è scambiabile ($\square$!) con

$\Phi(A \cup B) \geq \max[\Phi(A), \Phi(B)]$ oppure con
$\Phi(A \cap B) \leq \min[\Phi(A), \Phi(B)]$

Esercizio. La distribuzione *duale* $\Phi^*(A) \doteq 1-\Phi(\overline{A})$ è anch'essa una distribuzione di fiducia. La dualità è *involutoria*: $(\Phi^*)^* = \Phi$. La distribuzione $\Phi = \lambda\Phi_1 + \overline{\lambda}\Phi_2$, combinazione convessa delle due distribuzioni di fiducia Φ_1 e Φ_2, è anch'essa una distribuzione di fiducia. Le probabilità bayesiane sono *autoduali*: $\Phi^* = \Phi$.

Il lettore obietterà a ragione che le distribuzioni di fiducia sono un concetto astratto, ma noi le adopererememo solo come "cornice di riferimento". Nella letteratura sono chiamate, ciò che tuttavia non è molto euristico, *capacità di Choquet* o *misure sfocate*, dove il temine "misura" viene usato nel senso dell'analisi; per noi la misura di una distribuzione è uno scalare, come nel caso della "misura" dell'incertezza H(P).

Ci interessano i due casi limite, quando le disuguaglianze sono uguaglianze: si ottengono, rispettivamente, le distribuzioni di *possibilità* e di *necessità*, che sono concetti di tipo logico.

Assiomi delle distribuzioni di possibilità:
$\Pi(\emptyset) = 0, \Pi(\mathcal{X}) = 1$
$\Pi(A \cup B) = \max[\Pi(A), \Pi(B)]$

Assiomi delle distribuzioni di necessità:
$N(\emptyset) = 0, N(\mathcal{X}) = 1$
$N(A \cap B) = \min[N(A), N(B)]$

La "filosofia" di fondo (che approfondiremo quando parleremo di "attestazioni consonanti") all'ingrosso suona: un'evento è possibile sse è possibile almeno uno degli eventi elementari che lo costituiscono, è necessario sse sono necessari tutti gli eventi elementari che lo costituiscono. In effetti non abbiamo *due* teorie nuove, ma una sola: possibilità e necessità sono infatti concetti *duali*. Data una distribuzione di possibilità si ottiene subito una distribuzione di necessità, e viceversa, tramite la relazione (□ !):

$$\Pi(A) + N(\overline{A}) = 1$$

Teorema debole: $N(A) \leq \Pi(A)$

Dimostrazione. Dobbiamo dimostrare $\Pi(A) + \Pi(\overline{A}) \geq 1$; ciò è vero poichè
$1 = \Pi(A \cup \overline{A}) = \max[\Pi(A), \Pi(\overline{A})]$.

D'ora in avanti, salvo avviso specifico, parlando di possibilità e di necessità daremo per scontata la loro dualità. Il teorema consente di associare all'evento A l'intervallo non-vuoto [necessità, possibilità]:

$$A \to [N(A), \Pi(A)]$$

Il teorema debole si può rafforzare:

Teorema forte: L'intervallo $[N(A), \Pi(A)]$ è di uno dei tipi che seguono: $[0,0]$ (impossibilità totale), $[1,1]$ (necessità totale, certezza totale), $[0,1]$ (ignoranza totale), $[0,\alpha]$, $[\alpha,1]$ (contingenza); rimane escluso il tipo $[\alpha,\beta]$ con $0 < \alpha \leq \beta < 1$.

Dimostrazione. Poichè $1 = \Pi(A \cup \overline{A}) = \max[\Pi(A), \Pi(\overline{A})]$, almeno uno dei due eventi A e $\overline{A}$ è pienamente possibile (ha possibilità 1).
Dualmente, poichè $0 = N(A \cap \overline{A}) = \min[N(A), N(\overline{A})]$, almeno uno è pienamente non necessario (ha necessità zero).
Dunque:
$\Pi(A) < 1$ implica $\Pi(\overline{A}) = 1$, ossia $N(A) = 0$;
$N(A) > 0$ implica $N(\overline{A}) = 0$, ossia $\Pi(A) = 1$. $\square$

In pratica, le possibilità e le necessità si assegnano mediante un vettore non negativo Π, come segue:

$$\mathcal{X} = \{x_1, x_2, ..., x_K\}$$
$$\Pi = (\pi_1, \pi_2, ..., \pi_K) \text{ dove } \max_i \pi_i = 1$$
$$A \subseteq X, \ \ \Pi(A) = \max_A \pi_i, \ \ N(A) = 1 - \Pi(\overline{A})$$

La formula $\Pi(A) = \max_A \pi_i$ esprime la "filosofia di fondo" della teoria; di norma si parte dalle possibilità e poi si *calcolano* le necessità. Dualmente, potremmo anche partire dal vettore delle necessità dei K co-singoletti (*sic*), ma ciò sarebbe meno naturale. Attenzione: la necessità di A *non* è uguale alla minima necessità dei singoletti che lo compongono, bensì alla minima necessità dei co-singoletti degli elementi che *non* lo compongono ($\square$, usa la dualità; il co-singoletto di $\{x\}$ è composto da tutti gli elementi dell'*universo di riferimento* (dello spazio campionario), *meno a*).

Esempio. L'unica informazione che si ha è che "solo i singoletti x_i di F sono possibili, gli altri no", e non si entra in ulteriori dettagli. Si tratta di uno stato di conoscenza descritto tramite un unico *insieme focale*. Poni allora $\Pi(x_i) = 1$ oppure 0, a seconda che x_i stia in F o non ci stia; gli intervalli $[N(A), \Pi(A)]$ sono dei tre tipi $[0,0]$ se $A \cap F = \emptyset$, $[1,1]$ se $A \subseteq F$, $[0,1]$ altrimenti.

Esempio delle uova:
Per chiarire la differenza fra i concetti di probabilità e di possibilità faremo ricorso a un esempio di Zadeh. Si pensi all'evento incerto "Domani mattina Pierino mangerà x uova a colazione", dove x può essere $0, 1, 2, 3, 4, \geq 5$. Nella seconda riga della matrice che segue ci sono le possibilità, nella terza le probabilità.

0	1	2	3	4	≥ 5
1	1	.9	.5	.1	0
.1	.8	.1	0	0	0

E' del tutto possibile che Pierino mangi due uova, ma un'accurata indagine statistica ha mostrato che lo fa solo il 10 per cento delle volte: $\text{Prob}\{x = 2\} = 0,1$; la valutazione possibilistica ha natura logica, quella probabilistica ha natura empirica. Sottolineiamo che la possibilità può essere molto più elevata della probabilità: nel nostro caso Poss$\{x = 0\} = 1$ mentre Prob$\{x = 0\} = 0,1$. Il fatto che la possibilità di "x=4" sia 0,1 può essere visto come una traduzione numerica dell'affermazione qualitativa "E' quasi impossibile che Pierino ce la faccia a mangiare quattro uova a colazione".

Nella gestione dell'incertezza si è spesso costretti a "dare i numeri". L'etica scientifica tradizionale (oggettiva) può sentirsi scossa, ma per consolarsi si rifletta che a procedure di questo genere si ricorre da tempo immemorabile in tutte le istituzioni scolastiche nelle quali si assegnano voti numerici.
I risultati statistici di una serie di esperimenti costituiscono uno stato di conoscenza per il soggetto che ne viene a conoscenza, e dunque nulla vieta, come si è fatto nel nostro esempio, di usare le probabilità oggettive, statistiche, anche come probabilità soggettive, epistemiche; in questo senso le probabilità soggettive "includono" le probabilità oggettive.

Il *principio di coerenza possibilità-probabilità* impone di scegliere $P(A) \leq \Pi(A)$; non è un'implicazione formale, ma una richiesta di contenuto, cui è "ragionevole" attenersi.

Esercizio: $\forall A \;\; P(A) \leq \Pi(A)$ implica $\forall A \;\; P(A) \geq N(A)$, e viceversa.

Esercizio: un'assegnazione congiunta di probabilità e possibilità non è sempre coerente se è $p_i \leq \pi_i$ soltanto per i K singoletti; esempio: $P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $\Pi = (1, \frac{1}{4}, \frac{1}{4})$. Bastano tuttavia K controlli, come mostra il criterio che segue. *Criterio*: ipotizzando che l'alfabeto sia ordinato per possibilità decrescenti (non crescenti), l'assegnazione congiunta è coerente se e solo se $\pi_i \geq P(A_i)$ per tutti i *segmenti finali* $A_i = \{a_i, a_{i+1}, a_{i+2}, \ldots, a_{K-1}, a_K\}$ (a esser pignoli bastano $K-1$ controlli, perchè il primo ha esito scontato ...).

Può succedere che, viste come distribuzioni di fiducia, una distribuzione di probabilità bayesiane e una distribuzione di possibilità coincidano, nel senso che assegnano la stessa fiducia a ciascun evento. E' importante saper confrontare la "capacità rappresentativa" di due teorie alternative di gestione dell'incertezza.

Il seguente teorema mostra che l'intersezione fra le due teorie, probabilità e possibilità, è "quasi vuota". Una distribuzione di fiducia si dice *deterministica su* x se è bayesiana deterministica con Prob$\{x\} = 1$.

Teorema. Una distribuzione di probabilità P e una distribuzione di possibilità Π coincidono se e solo se sono deterministiche su x. Una distribuzione di probabilità P e una distribuzione di necessità N coincidono se e solo se sono deterministiche su x. Una distribuzione di possibilità Π e una distribuzione di necessità N (duali o non duali) coincidono se e solo se sono deterministiche su x.

Dimostrazione. Che il determinismo su x della distribuzione di fiducia Φ implichi $\Phi = P = \Pi = N$ è ovvio. Passiamo al viceversa; la seconda affermazione scende dalla prima per dualità , poiché le probabilità bayesiane sono *autoduali*: $P^*(a) \doteq 1 - P(\overline{A}) = P(A)$.

Se $\forall A, P(A) = \Pi(A)$, si ha $P(\mathcal{X}) = \Pi(\mathcal{X})$, ossia $\sum p_i = 1$, e anche $\max p_i = 1$, per cui un solo addendo p_i è non nullo, e vale proprio 1.
Se $\forall A, N(A) = \Pi(A)$ si ha $\forall x, N(x) = \Pi(x)$. Ora, al più un singoletto può avere necessità positiva ($\Box$! se fossero due, intersecandoli ...).

2. Insiemi sfocati e logiche multivalenti

In questa sezione più che di probabilità eterodosse parleremo di *insiemi* eterodossi.

Un sottinsieme ordinario $A \subseteq \mathcal{X}$ (sottinsieme *nitido*, *crisp subset*), può venir "identificato" con la sua *funzione caratteristica* $f_A : \mathcal{X} \to \{0,1\}$. Per gestire la nozione di "vaghezza" Lotfi Zadeh introdusse nel 1965 gli *insiemi sfocati* (o *sfumati*, *fuzzy sets*, *ensembles flous* in francese, ossia vaporosi, pastosi, indistinti), consentendo che la funzione caratteristica assumesse i propri valori su *tutto* l'intervallo $[0,1]$, per cui un'elemento x_i può appartenere al sottinsieme solo "in una certa misura". Il numero $f_A(x_i)$ è appunto il *grado di appartenenza* dell'oggetto x_i all'insieme sfocato A. L'identificazione fra insieme sfocato e funzione caratteristica (*funzione di appartenenza*, *membership function*) è così stretta che scriveremo concisamente $A(x_i)$ invece di $f_A(x_i)$.

Dilemmi docimologici (lettura). Pierino va in terza elementare e quest'anno sperimenterà ancora una volta un nuovo sistema di valutazione. Non più le lettere (che in seconda "scendevano" dalla A alla E: A > B > C > D > E, se ci è consentito il simbolismo matematico), ma, più umanamente, dei giudizi: ottimo, buono, discreto, sufficiente e insufficiente (l'ordinamento è scontato; a proposito, sarebbe discretamente > benino, benino > discretamente oppure benino = discretamente ?). Per inciso: qualche anno fa in un istituto universitario, che per la cronaca era del versante umanistico, si decise di sostituire ai freddi voti in trentesimi dei giudizi linguistici, del tipo di quelli che si ritroverà a ricevere Pierino. Alla valutazione finale, nella quale si richiede un consuntivo unico, si incappò nel problema della media: quanto fa ottimo + buono + ottimo diviso 3, per esempio? Si usci dalle secche grazie a una codifica numerica del tipo ottimo = 30, buono = 28, ecc., per cui alla domanda precedente si risponde ventinove virgola tre periodico, un numero che ha l'aria di essere particolarmente preciso, avendo un'infinità di cifre decimali; a molti venne il dubbio di essere incappati in un deplorevole circolo vizioso. Di recente un insegnante liceale (di matematica, con nostro sconcerto) si è lamentato che rinunciando ai numeri si siano persi appunto i vantaggi della precisione: un tempo era chiarissimo che se Tizio prendeva 7 nel tema di latino, Caio prendeva 8, e Sempronio 7/8 (curiosa notazione scolastica per indicare sette e mezzo e non sette ottavi), allora il tema di Sempronio stava esattamente a metà strada fra quello di Tizio e quello di Caio. Con maggior estremismo numerico si potrebbe sostenere che l'otto in ginnastica sta esattamente a metà strada fra il sette in disegno e il nove in quella che una volta si chiamava buona condotta. Anche se in forma codificata, è dalla notte dei tempi che i docenti danno i numeri; se il gioco di parole sottolinea l'inevitabile arbitrarietà dei giudizi, il senso di responsabilità sociale ci fa capire che altrettanto inevitabile è il ricorso a queste opinabili procedure. Tuttavia le ambasce del Ministero della Pubblica Istruzione sono emblematiche di un tormento epistemico molto più esteso. Il numero serve per contare e per misurare. Il primo uso, che si accontenta dei numeri interi, 1, 2, 3 ecc., magari col segno meno per denotare mancanza, è abbastanza tranquillo: tranne al più accanito dei sofisti è inequivocabile che cosa intendiamo dicendo di possedere tre vacche, mentre il vicino ne possiede cinque, anche se basta che l'inverno scorso uno si sia buscato tre raffreddori, a differenza del vicino che se n'è buscati cinque, per cominciare ad avere qualche grana semantica. Ma quando si passa al numero che misura, il confine fra uso legittimo e uso illegittimo diventa sfumato e opinabile. Che cosa si può misurare? La superficie di questa stanza, d'accordo, o la sua temperatura, salvo esser disposti a ignorare la differenza fra la temperatura misurata vicino al pavimento, al soffitto o alla finestra; ma è lecito misurare tramite numeri l'attendibilità di un testimonio in tribunale, e affermare che l'attendibilità di Tizio è un terzo di quella di Caio? Il numero, nostro servitore e nostro tiranno, aiuta o depista? In termini che qualcuno riterrà fuori moda: qual è il confine fra l'uso scientifico e quello non-scientifico del numero (fra oggettivo e soggettivo, e poi fra soggettivo e arbitrario)? Questo confine si è spostato nel tempo, proprio perchè la nostra concezione della scientificità è soggetta alle vicende e alle trasformazioni storiche. Gli insegnanti e i pedagoghi possono mettersi il cuore in pace: oggidì anche gli scienziati e i tecnologi si sono piegati a "dare i numeri", a riprova della vaghezza del confine che separa scienza e non-scienza. Assicurano gli storici che la logica binaria, in bianco e nero, la logica che non conosce le sfumature di grigio, a torto viene attribuita ad Aristotele; si è suggerito pertanto di parlare di logica crisippina e non di logica aristotelica, la quale è ternaria, poichè conosce anche il valore logico "possibile", intermedio fra vero e falso, e che compete ad esempio alla frase "domani pioverà". Le logiche a molti valori logici (perfino a infiniti valori diffusi sull'intervallo che va da 0 = falso a 1 = vero) sono fiorite nel nostro secolo e sono legate soprattutto al nome del logico polacco Jan Łukasiewicz. Quando però si è trattato di simulare in un sistema esperto il ragionamento umano e le sue efficacissime incongruenze neppure questo è bastato; si è dovuto ricorrere anche a logiche non-monotòne: esse ammettono la ritrattazione, come nel sillogismo "gli uccelli volano", "Titti è un uccello", dunque "Titti vola", in cui la conclusione va ritrattata quando si scopra che Titti è uno struzzo. Torniamo alle logiche multivalenti e alla loro versione di maggior successo, quella che le lega agli insiemi sfocati (*fuzzy sets*) proposti, dapprima con scarso successo, da Lotfi Zadeh nel 1965. Ancora più immediato è qui il legame con gli insegnanti in classe, tanto che quotidiane situazioni scolastiche servono benissimo a introdurre certi concetti delle logiche sfocate. I voti scalati fra 0 e 1 invece che fra 0 e 10 sono interpretabili come valori logici sfumati, sicchè se lo studente S avesse ricevuto 7 in latino, il logico fuzzy, o comunque qualche logico fuzzy, direbbe che il valore logico della proposizione "S conosce il latino" è appunto 0,7 (chi alla logica preferisce gli insiemi direbbe invece che il grado di appartenenza di S all'insieme sfocato degli studenti che sanno il latino è 0,7). Le logiche e gli insiemi fuzzy hanno avuto all'inizio vita difficile, colpevole forse anche il nome fuzzy, sfocato, poco attraente sia per lo scienziato tradizionale che per il tecnologo alla caccia di applicazioni redditizie. (La connotazione negativa viene ribaltata nella traduzione italiana di insieme sfumato; con maggior coraggio in francese e in spagnolo si dice *ensemble flou* e *conjunto borroso*; i tedeschi si sono accontentati di *Fuzzymengen*.) Sono stati i giapponesi che hanno adoperato la teoria sfocata nelle applicazioni industriali, con un successo enorme, ciò che ha ben presto riacceso l'interesse degli occidentali, in misura tale che oggi la qualifica fuzzy (non tradotta e probabilmente arbitraria) si può trovare nella pubblicità spicciola di lavatrici e cineprese, le cui magiche prestazioni sarebbero assicurate dal fatto di essere provviste di un'intelligenza fuzzy (in Giappone il termine inglese, a quanto sembra, è entrato nell'uso popolare). Se il termine venisse tradotto anche i massai e le massaie italiane, come quelli inglesi e americani, potrebbero apprezzare lo splendido ossimoro (tecno)logico delle macchine fotografiche sfocate che fanno per ciò stesso fotografie nitidissime. Fuzzy is crisp è uno slogan che richiama certe incredibili associazioni degli opposti tipiche delle culture orientali, a partire dalla polarità fra lo yin femminile e lo yang maschile: e i sociologi hanno speculato sull'interesse dell'oriente per una teoria nata in occidente (peraltro lo statunitense Zadeh è in origine un persiano dell'Azerbaigian allora sovietico), ma forse più confacente alla mentalità orientale. Sia come sia, i successi tecnologici ottenuti nell'estremo oriente, e la loro tangibile controparte in termini di profitti, hanno riaperto, anzi spalancato, le porte delle accademie occidentali alle logiche sfumate, con il che la sfera d'azione del numero si è ancora dilatata, e le frontiere della scientificità hanno dovuto concedere nuovo spazio alla soggettività. Il professore di matematica che ha dato voti numerici da sempre può ormai farlo senza dover violare la sua etica professionale. Torniamo al dilemma di partenza. Perchè il numero? Perchè non accontentarsi di ordinamenti solo qualitatitivi, come fa la lingua naturale con cui ci esprimiamo nella cosiddetta vita di ogni giorno, visto che è la lingua naturale a muovere quella splendida e insieme carente "macchina inferenziale" che è il cervello umano? Perchè la camicia di forza degli ordinamenti lineari, mentre la lingua ci offre la flessibilità di ordinamenti ramificati, a livelli sovrapposti e insieme annidati, di una topologia ben difficilmente dominabile? Prendiamo il caso dell'insieme sfocato degli oggetti verdi presenti in un certo laghetto, cui appartengono, ma in grado diverso, sia le rane sia le ninfee. Come si fa a dare i numeri? Verdastro è meno verde di verde: verdastro $<$ verde nell'ordinamento per verdità, ma se passiamo ai numeri su scala zero - uno, come fanno i "fuzzisti", al verde spetta 1, certo, ma a verdastro quanto spetta, 1/2, 1/4, 3/4? e se verde $= 1$, che cosa facciamo di verdissimo, che pure nella lingua si trova? e se a giallo spetta zero nell'ordinamento per bluità, che cosa spetta a verde, che è più blu di giallo, come sanno quelli cui piace mescolare i colori? Eppure senza il numero ci si incaglia nelle stesse difficoltà, anzi rafforzate, che trova l'insegnante in classe. Certe potenzialità del numero hanno un fascino a cui è difficile resistere. Stabiliti i due estremi (falso e vero, che nulla impedisce di chiamare 0 e 1) è naturale pretendere che ci sia un punto di indifferenza nel quale "non si sa che pesci pigliare", e che non si può non codificare tramite il valore numerico un mezzo; ma allora perchè rinunciare a un punto di indifferenza intermedio fra falso e un mezzo, ossia a un quarto, e fra un mezzo e uno, ossia a tre quarti? E' irrestibile il desiderio di poter confrontare incrementi (ad esempio l'aumentata efficienza degli allievi di una classe) e non solo valori: ma "buono meno discreto" è più grande o più piccolo di "discreto meno sufficiente"? Abbiamo già ricordato il problema delle medie, ossia degli indici sintetici. Per queste e per altre ragioni il cervello artificiale non numerico prima o dopo si incastra e diventa poco efficiente. Il successo sul piano dell'efficienza spicciola se non su quello della *finesse* è nei dati di fatto: le diagnosi mediche automatiche, per esempio. Esse consentono o consentiranno tassi di errore più bassi di quelli raggiunti dall'esperto umano, anche se pongono sul tappeto tutta una serie di problemi morali e giuridici ineludibili. (Chi è responsabile di una diagnosi automatica sbagliata? Che cosa rimpiazzerà i meriti terapeutici del cosiddetto "rapporto umano" in un mondo in cui la solitudine diventa sempre di più una grave malattia sociale?)

3. L'algebra degli insiemi sfocati

L'algebra degli insiemi sfocati è retta dalle definizioni che seguono di unione, intersezione, complementazione e inclusione:

$$\begin{aligned}
&[A \cup B](x_i) = \max[A(x_i), B(x_i)]\,, \quad \text{ossia } f_{A\cup B}(x_i) = \max[f_A(x_i), f_B(x_i)] \\
&[A \cap B](x_i) = \min[A(x_i), B(x_i)] \\
&\overline{A}(x_i) = 1 - A(x_i) \\
&A \subseteq B \iff \forall x_i \; A(x_i) \leq B(x_i)
\end{aligned}$$

Si ritrovano le regole abituali nel caso dei (sott)insiemi nitidi, per i quali la funzione caratteristica assume solo i valori 0 e 1. Segue un elenco di proprietà facilmente verificabili:

commutatività e associatività : $A \cup B = B \cup A\,,\; A \cup (B \cup C) = (A \cup B) \cup C$
idempotenza e involuzione : $A \cup A = A\,,\; A = \overline{\overline{A}}$
distributività : $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
legge di De Morgan : $\overline{A \cap B} = \overline{A} \cup \overline{B}$

oltre a quelle *duali* che si ottengono scambiando fra di loro i segni di unione e di intersezione, in particolare l'altra legge di De Morgan: $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

Sottolineamo che *non* valgone le due regole "nitide":

$$A \cap \overline{A} = \emptyset \;\text{ e }\; A \cup \overline{A} = \mathcal{X}$$

Non valgono dunque nè il principio di *non contraddizione*, nè il principio del *terzo escluso*. Tutto ciò che si può dire è :

$$A \cap \overline{A} \subseteq F \subseteq A \cup \overline{A}$$

dove F è l'insieme *totalmente sfocato*, la cui funzione caratteristica ha il valore costante $\frac{1}{2}$.

Agli insiemi sfocati è legata la *logica sfocata*, nella quale non valgono appunto nè il principio di non contraddizione nè quello del terzo escluso: gli strumenti sfocati hanno portato a straordinari successi pratici nella gestione "intelligente" dell'incertezza mediante il *controllo sfocato*. Il legame fra insiemi e logica appare chiaro appena si 'identifichino" sottinsiemi e proprietà: ad esempio il sottinsieme dei numeri pari viene identificato con la proprietà "essere un numero pari", per cui appartenere al sottinsieme equivale a verificare la proprietà. In ambiente sfocato si "confondono" le due affermazioni:
Il grado di appartenenza dell'oggetto x *all' insieme* A è μ.
Il valore logico (il valore di verità) della proposizione "x *verifica la proprietà* A" è μ.
Alle operazioni insiemistiche di unione, intersezione e complementazione corrispondono le operazioni logiche di disgiunzione (OR, oppure), congiunzione (AND, e anche) e negazione (NOT), all'inclusione insiemistica corrisponde l'implicazione logica. Ciò sia nel caso nitido sia in quello sfocato.

A un insieme sfocato A di funzione caratteristica $A(x)$ vengono associati i seguenti insiemi nitidi (α è un numero reale dell'intervallo $[0,1]$):

$$\textit{supporto di } A = \{x : A(x) > 0\}$$

$$\textit{sezione o taglio di livello } \alpha = \{x : A(x) \geq \alpha\}$$

Ovviamente, al crescere di α, l'α-taglio può solo rimpicciolire: $\alpha < \beta$ implica $A_\beta \subseteq A_\alpha$. Oltre agli α-tagli *deboli*, come quelli che abbiamo introdotto, si considerano talvolta gli α-tagli *forti*, definiti tramite una disuguaglianza stretta; il supporto coincide con lo 0-taglio forte. Gli α-tagli (*α-sezioni*, *α-cuts*) danno un primo esempio di *messa a fuoco* (*focalizzazione*, *defuzzification*): si fissa una soglia α, vengono "bocciati" tutti gli oggetti con grado di appartenenza insufficiente, mentre vengono promossi quelli con il grado di appartenenza sufficiente, pari almeno ad α. Il prezzo della messa a fuoco, che "forza" l'insieme sfocato alla nitidezza, è una perdita di flessibilità, e dunque una perdita di informazione.

Un sottinsieme sfocato si dice *normale* se c'è almeno un elemento che gli appartiene "al cento per cento" (in esso la funzione caratteristica vale 1), ossia se il suo *nucleo* non è vuoto (il nucleo è l'α-taglio di livello 1), ossia ancora se è uguale a 1 la sua *altezza* (l'altezza di un insieme sfocato è il massimo dei suoi gradi di appartenenza).

C'è una stretta parentela fra la teoria della possibilità e quella degli insiemi sfocati normali (o meglio dei *sottinsiemi* sfocati normali; l'insieme ambiente $\mathcal{X}$ è comunque nitido). La si mette in luce "confondendo" le due affermazioni:
Il grado di appartenenza dell'oggetto x *all' insieme* A *è* μ.
La possibilità che l'oggetto x *goda della proprietà* A *è* μ.

3 bis. L'implicazione sfocata.

Riprendiamo la stretta parentela che c'è fra la teoria degli insiemi sfocati e la logica multivalente; la si mette in luce "confondendo" le due affermazioni:
Il grado di appartenenza dell'oggetto x *all' insieme* A *è* μ.
Il valore logico della proposizione $x \in A$ *(il valore logico della proposizione: l'oggetto* x *gode della proprietà* A*) è* μ.

Nella logica binaria o crisippina l'implicazione fra due proposizioni $A \to B$ viene "identificata" con la disgiunzione $\overline{A} \vee B$, con il risultato -un pochino umoristico anche se il senso dell'umorismo dei logici classici in questo frangente evapora- che l'implicazione "se mia nonna avesse le ruote sarebbe un tram" si becca il valore logico 1 (se l'*antecedente* è falso, qualunque sia il *conseguente*, l'implicazione è comunque vera). Si può scimmiottare la logica classica arrivando all'implicazione fuzzy chiamata nella letteratura *implicazione di Kleene-Dienes* o *di Zadeh*:

$$[A \to B](x) \;=\; \max[1 - A(x), B(x)]$$

A dire il vero, quando si passò alle applicazioni industriali della logica fuzzy, le nonne rotanti si rivelarono una gran fonte di guai: a differenza di quanto accade nell'empireo logico, nel mondo reale ci si può fidare solo di implicazioni in cui l'antecedente è vero o quasi, per cui i pratici e i tecnologi ricorrono di solito alla cosiddetta *implicazione di Mamdani*, che coincide pari pari con la congiunzione:

$$[A \to_M B](x) \;=\; \min[A(x), B(x)] \;\text{ ossia }\; A \to_M B = A \wedge B$$

Il lettore non si scandalizzi se in letteratura c'è più di un'implicazione: pur di rado, si usano perfino altre congiunzioni ed altre disgiunzioni. Per rispetto dei padri fondatori, menzioneremo la *digiunzione* e la *congiunzione di Lukasiewicz*:

$$[A \vee_L B](x) \;=\; \min[A(x) + B(x), 1]\;, \;\; [A \wedge_L B] \;=\; \max[A(x) + B(x) - 1, 0]$$

con la rispettiva (e celebre) *implicazione*:

$$[A \to_L B](x) \;=\; \min[1 - A(x) + B(x), 1] \;\text{ ossia }\; A \to_L B = \overline{A} \vee_L B$$

Al lettore il compito di verificare che le formule di De Morgan si estendono alle operazioni alla Lukasiewicz. Esiste tutta una teoria matematica delle operazioni logiche astratte, di cui quelle alla Zadeh o alla Luksiewicz sono casi speciali; in questa teoria le congiunzioni e le disgiunzioni generalizzate vengono chiamate, rispettivanmente, *t-norme* e *t-conorme*.

4. Quantità sfocate.

In questa sezione, eccezionalmente, l'ambiente di riferimento non è più finito; le nozioni della sezione precedentemente sono tuttavia generalizzabili senza difficoltà. Se l'insieme ambiente è l'insieme $\mathbf{R}$ dei numeri reali, un sottinsieme sfocato A si chiama anche *quantità sfocata* (*fuzzy quantity*); la sua funzione caratteristica è dunque un'applicazione $A(x)$ che mappa $\mathbf{R}$ nell'intervallo $[0, 1]$. Attenzione: *formalmente* quantità reali sfocate e sottinsiemi sfocati di $\mathbf{R}$ sono la stesssa cosa, ma il modo di "pensarli" e di "vederli" *non* lo è affatto! (Ciò apparirà molto chiaro quando più oltre la possibile confusione fra inclusioni insiemistiche e ordinamenti numerici ci creerà qualche grana.)

Esempio. Fra le quantità sfocate si segnalano i *numeri sfocati triangolari.* Il loro supporto è un intervallo $[a, b]$; il grado di appartenenza 1 viene raggiunto in un solo valore r dell'intervallo; la funzione di appartenenza $A(x)$ sale linearmente da 0 a 1 sul sotto-intervallo $[a, r]$ e discende sempre linearmente da 1 a 0 sul restante sotto-intervallo $[r, b]$. Il numero triangolare si indica spesso con la terna (a, r, b), che è sufficiente a identificarlo; si usano scritture del tipo di $\tilde{r}$, da leggere: circa r. Non rimangono esclusi i casi in cui ci siano coincidenze fra a, r e b; in particolare il numero sfocato (r, r, r), la cui funzione caratteristica vale 1 in corrispondenza a r ed è nulla altrove, si può confondere con il "normale" numero reale nitido r.

Esempio. Popolari quantità dell'artitmetica sfocata sono gli *intervalli trapezoidali* (a, r, s, b), per i quali la funzione di appartenza è appunto un trapezio; $[a, b]$ è il supporto, $[r, s]$ è il nucleo. I normali intervalli nitidi (chiusi) si ritrovano quando nucleo e supporto coincidono. Per $a \le r = s \le b$ si ritrovano i numeri sfocati triangolari.

Per poter effettuare operazioni aritmetiche su quantità sfocate in maniera sensata, dobbiamo prima introdurre un importante principio della teoria che si chiama il *principio di estensione* (*extension principle*): esso consente di estendere “meccanicamente” molte nozioni nitide all'ambiente sfocato.

Rimaniamo per un istante in ambiente nitido. Se $f : \mathcal{X} \to \mathcal{Y}$ è una funzione da un insieme-input finito $\mathcal{X}$ a un insieme-output $\mathcal{Y}$, rimane automaticamente definita la sua estensione $f : 2^{\mathcal{X}} \to 2^{\mathcal{Y}}$ dall'insieme delle parti nitide di $\mathcal{X}$ all'insieme delle parti nitide di $\mathcal{Y}$. Basta porre per $A \subseteq \mathcal{X}$:

$$f(A) = \{y : \exists x \in A \text{ per cui } f(x) = y\} \subseteq \mathcal{Y}$$

A parole, ipotizzando solo per esprimerci meglio che l'insieme $\{x : f(x) = y\} \equiv \{x_1, x_2, \ldots, x_s\}$ sia finito: il grado di appartenenza di y ad $f(A)$ è 1 se almeno una delle seguenti affermazioni è vera (ha grado di verità 1): $x_1 \in A$ oppure $x_2 \in A$ oppure ... oppure $x_s \in A$. Altrimenti detto: il valore logico di $f(A)$ in y è il valore logico di $A(x_1) \vee A(x_2) \vee \ldots \vee A(x_s)$. Conta dunque il *massimo* dei gradi di appartenenza $A(x_1)$, $A(x_2)$, ..., $A(x_s)$ (il lettore apprezzerà ancora una volta l'analogia fra l'operazione logica di disgiunzione = oppure = OR, e l'operazione aritmetica di massimo). Dunque:

$$B = f(A) \iff^{def} B(y) = \sup_{x: f(x)=y} A(x)$$

dove abbiamo scritto estremo superiore *sup* invece di massimo perchè l'insieme $\{x : f(x) = y\}$ potrebbe ben essere infinito. Ebbene, l'ultima formula funziona pari pari anche se A è un sottinsieme sfocato; anche l'insieme immagine $B = f(A)$ è sfocato. Il principio di estensione è appunto la formula appena scritta, e consente estendere la funzione f a una funzione dall'insieme delle parti *sfocate* di $\mathcal{A}$ all'insieme delle parti *sfocate* di $\mathcal{B}$. In altri termini la funzione f nasce come trasformazione fra numeri reali nitidi, ma grazie al principio di estensione cresce, e diventa una funzione che a quantità sfocate associa altre quantità sfocate. Una nota scontata ma doverosa: se l'insieme $\{x : f(x) = y\}$ su cui si calcola l'estremo superiore è vuoto, si pone $\big[f(A)\big](y) = 0$; è un'osservazione che useremo fra poco nel calcolo degli inversi e dei quadrati sfocati.

Esercizio. Se A è una quantità sfocata di funzione caratteristica $A(x)$, calcola la funzione caratteristica delle quantità sfocate $-A$ e $|A|$ (l'opposto e il valore assoluto di A). Si tratta di applicare ad A il principio di estensione con le funzioni $y = f(x) = -x$ e $y = g(x) = |x|$, rispettivamente. Si trova $\big[-A\big](y) = A(-y)$, $|A|(y) = \max[A(-y), A(y)]$ per $y \geq 0$, altrimenti $|A|(y) = 0$.

Esercizio. Se f è biunivoca (*one-to-one*, invertibile alla f^{-1}), si ha

$$B(y) = A\big(f^{-1}(y)\big)$$

Esercizio. Calcoliamo l'inverso $B = 1/A$; stavolta $f(x) = 1/x$. Si ha $B(y) = 0$ per $y = 0$ (il campo d'azione del *sup* è vuoto); stante la biunivocità di f, si ha $B(y) = A(1/y)$ per $y \neq 0$. Tanto per fissare le idee, supponiamo che A sia il numero triangolare $(-1, 0, +1)$.

Allora $B(y) = 1 + 1/y$ per $y \leq -1$, $B(y) = 0$ per $-1 \leq y \leq 1$ e $B(y) = 1 - 1/y$ per $y \geq 1$; dunque la funzione di appartenenza $B(y)$ dell'inverso $1/A$ decresce da 1 a 0 sulla semiretta $y \leq -1$, è nulla sull'intervallo $[-1, +1]$ e ricresce a 1 sulla semiretta $y \geq 1$; sulle due semirette la funzione di appartenenza $B(y)$ è concava (*convex-cap*, concavità verso il basso).

Esercizio. Calcoliamo il quadrato $B = A^2$; $f(x) = x^2$. Si ha $B(y) = 0$ per $y \leq 0$; per $y \leq 0$, $B(y) = \max\big[A(-\sqrt{y}), A(+\sqrt{y})\big]$; $\square$. Ad esempio, se A è il numero triangolare $(0, 1, 2)$, la funzione di appartenenza $B(y)$ è nulla sul semiasse negativo e sulla semiretta $y \geq 4$, cresce da 0 a 1 sull'intervallo $[0, 1]$ (funzione concava), e poi ridecresce a 0 per y che va da 1 a 4 (funzione convessa, *convex-cup*). Si noti che il risultato *non* è triangolare. Al lettore calcolare il quadrato di $(-1, 0, +1)$.

Una generica operazione aritmetica $\bullet$ fra due argomenti x e y ha la forma $z = x \bullet y$; vogliamo estenderla a un operazione fra quantità sfocate $C = A \bullet B$. Nel principio di estensione l'insieme su cui si calcola l'estremo superiore per ottenere $[A \bullet B](z)$ è $\{x, y : x \bullet y = z\}$; non è però chiaro quale sia il grado di appartenenza della *coppia* x, y all'insieme delle coppie $A \times B$. In altre parole dobbiamo ancora definire in maniera sensata il *prodotto cartesiano* (insieme delle coppie) di due insiemi sfocati. Nel caso nitido una coppia x, y appartiene all'insieme prodotto $A \times B$ quando si ha *sia* $x \in A$ per la prima componente *sia* $B \in B$ per la seconda componente; all'operazione logica di congiunzione (*sia sia*) corrisponde l'operazione aritmetica di *minimo* fra i valori logici. Definiamo allora il prodotto cartesiano di insiemi sfocati ponendo:

$$[A \times B](x, y) \;=\; \min\big[A(x), B(y)\big] \;\equiv\; A(x) \wedge B(y)$$

dove, come si fa spesso, siamo giunti al punto di usare il simbolo logico $\wedge$ al posto del minimo; è una confusione in cui si incorre spesso perchè aiuta la leggibilità delle formule. Le operazioni aritmetiche fra due quantità sfocate, come la somma e il prodotto, si possono ora eseguire grazie al principio di estensione:

$$\big[A \bullet B\big](z) \;=\; \sup_{x, y : x \bullet y = z} A(x) \wedge B(y)$$

Segue una lista di proprietà di cui la somma, il prodotto e la divisione fra quantità sfocate fortunatamente godono; in essa intervengono i numeri nitidi 0 e 1 $\big($i numeri sfocati triangolari $(0, 0, 0)$ e $(1, 1, 1)\big)$. Sottolineiamo che le operazioni sfocate generalizzano quelle nitide, com'è doveroso.

$A + B = B + A \quad AB = BA$
$A + (B + C) = (A + B) + C \quad A(BC) = (AB)C$
$0 + A = A \quad 1A = A \quad 0A = 0$
$-(-A) = A \quad (-A)B = -(AB) = A(-B) \quad A + (-B) = A - B \quad \frac{A}{B} = A\frac{1}{B}$

La dimostrazione delle proprietà precedenti è facile e noiosa. Segue un'altra lista di proprietà di cui la somma, il prodotto e la divisione fra quantità sfocate sfortunatamente *non* godono:

$A(B+C) \neq AB+AC$
$A-A \neq 0 \quad (A+B)-B \neq A$
$\frac{A}{A} \neq 1 \quad \frac{A}{B}B \neq A$

Il disastro è mitigato dal fatto che la proprietà distributiva vale se il primo termine è nitido: $r(A+B) = rA + rB$. Fra le quantità sfocate sub-normali c'è quella, assai poco motivante, che corrisponde all'insieme *vuoto*, e che non ha nulla a che vedere con lo zero nitido (normale e triangolare, $a = r = b = 0$).

Esercizio. Da' (contro)esempi a sostegno delle mancate identità precedenti. Prendiamo il caso della mancata proprietà distributiva; si riesce a dimostrare soltanto:

$$\forall x \quad \big[A(B+C)\big](x) \leq \big[AB+AC\big](x)$$

ossia $A(B+C) \subseteq AB+AC$. Attenzione! La relazione precedente è un'inclusione fra insiemi: bisogna resistere alla tentazione di scrivere $A(B+C) \leq AB+AC$ con il simbolo aritmetico di $\leq$. Le nozioni sfocate devono estenderere quelle nitide, ma la relazione nitida $r \leq s$ non ha nulla da spartire con l'inclusione fra gli insiemi che corrispondono ai due numeri nitidi r ed s, ossia, se si vuole, ai due numeri triangolari (r,r,r) e (s,s,s). In altre parole: se valgono le inclusioni fra *insiemi* $A \subseteq B \subseteq \mathbf{R}$ non c'è nessuna ragione sensata di dire che la *quantità* sfocata A è più "piccola" della *quantità* sfocata B. In particolare non c'è nessuna ragione sensata di dire che la quantità sfocata $\mathbf{R}$, la cui funzione caratteristica è sempre uguale a 1 e che esprime uno stato di conoscenza nitido sì, ma di totale ignoranza su quale sia il numero reale di cui si sta parlando (tutto è possibile), sia la più "grande" quantità sfocata in circolazione. *Definire un ordinamento fra quantità sfocate, sia pur parziale, è assai problematico.*

Esercizio. Se $\tilde{r} = (a,r,b)$ e $\tilde{s} = (c,s,d)$ sono due numeri triangolari, la loro somma $\tilde{r} + \tilde{s}$ è il numero triangolare $(a+c, r+s, b+d)$. Questa comoda regola di calcolo contribuisce a spiegare il successo dei numeri triangolari nelle applicazioni della teoria sfocata; altrettanto comodo è sommare trapezi, inclusi trapezi "degeneri" con $s = b = +\infty$, che si usano volentieri nelle applicazioni. Purtroppo il prodotto di due numeri triangolari non è di norma triangolare; lo abbiamo già intuito calcolando il quadrato $\tilde{1}^2$ di $(0,1,2)$. *Nella matematica soft le complicazioni di calcolo sono mal tollerate*: è per questo che, limitatamente ai numeri triangolari a supporto positivo, il prodotto viene di solito sostituito da un più comodo "pseudoprodotto" in cui si "forza" il risultato alla triangolarità: $(a,b,c) * (\alpha,\beta,\gamma) = (a\alpha, b\beta, c\gamma),\ a, \alpha \geq 0$. Un'osservazione per i lettori più pazienti: purtroppo in generale $A^2 \neq A \times A$ (è per questo che qualche riga fa abbiamo scritto "intuito" invece di "dimostrato"); si puo' solo dimostrare che, per ogni x, $A^2(x) \leq [A \times A](x)$, il che naturalmente non implica affatto la disuguaglianza fra quantità sfocate $A^2 \leq A \times A$, ma solo l'inclusione fra insiemi sfocati $A^2 \subseteq A \times A$.

Abbiamo già parlato di *focalizzazione* a proposito degli α-tagli; nella teoria sfocata, tuttavia, il termine viene di solito riservato a una procedura diversa, che riguarda le quantità sfocate e che consiste nel trasformarle in un unico numero reale nitido, che in qualche modo le "riassume". La scelta più comune, che si ispira ai baricentri della fisica e alle speranze matematiche del calcolo delle probabilità, consiste nel ricorrere al *centroide* $C(A)$

della quantità sfocata A. Per attenersi all'analogia fisica, nel caso in cui il supporto della quantità sfocata sia un intervallo (o una semiretta, o l'asse reale) si pensa alla funzione caratteristica $A(x)$ come a una funzione di densità lineare che descrive una distribuzione di massa di tipo continuo. La formula del centroide è allora la stessa di quella del baricentro sul continuo (se il supporto fosse finito, gli integrali vanno sostituiti da somme; i più raffinati possono usare gli integrali di Lebesgue, che vanno bene comunque). Ci limiteremo a quantità sfocate a supporto intervallare $|a, b|$, in maniera da avere la garanzia che la "massa totale" $\int A(x)dx$ sia finita; ciò non è una restrizione sensibile nelle applicazioni pratiche che riguardano il *controllo sfumato*. A esser pignoli va aggiunta qualche ovvia condizione di regolarità su $A(x)$, che diamo per scontata, per garantire l'esistenza degli integrali; dobbiamo anche escludere che la "massa" totale sia nulla, ossia che l'insieme sfocato A sia vuoto.

$$C(A) \;=\; \frac{\int_a^b xA(x)dx}{\int_a^b A(x)dx}$$

Esercizio. Nota che il centroide di un numero sfocato $\tilde{r}$ non coincide sempre con r; nel caso triangolare, ad esempio, si ha coincidenza solo se $b - r = r - a$.

Esercizio. Un rivale del centroide, che viene anch'esso usato nel controllo sfumato, è il *centro dei massimi*. A parole è il punto di mezzo fra il più piccolo e il più grande x in cui la funzione caratteristica assume il suo valore massimo (in cui essa vale 1 se la quantità sfocata A è normale); al lettore il compito pignolo di scrivere la formula stando attento a usare *sup* e *inf* invece di *max* e *min*. Il centro dei massimi si definisce tal quale anche per quantità sfocate a supporto finito (purché non vuoto); il centro dei massimi di un numero sfocato $\tilde{r}$ è sempre r. Centroidi e centri dei massimi potrebbero forse venir usati per instaurare un ordinamento ragionevole fra quantità sfocate. Il lettore che sa di statistica riconoscerà nel centro dei massimi la *moda*; anche la *mediana* degli statistici potrebbe venir usata per focalizzare quantità sfocata: si tratta in questo caso di "spezzare a metà" la massa disponibile.

Complementi. Il principio di estensione ci permette di parlare di funzioni e operazioni sfocate in tutta generalità, ma spesso la nostra "intuizione aritmetica" viene messa a dura prova. Di fatto l'aritmetica fuzzy "che serve" e che non è "anti-intuituiva" è largamente triangolare e trapezoidale. Qui di seguito, tuttavia, introduciamo le versioni sfocate "generali" degli intervalli chiusi e dei numeri reali.

Definizione. Un *intervallo sfocato* è una quantità sfocata normale, il cui supporto è limitato e i cui α-tagli sono intervalli chiusi ($\alpha > 0$; ovviamente lo 0-taglio è sempre e comunque uguale a tutto l'asse reale).

Esercizio. Il supporto di un intervallo sfocato I è un intervallo nitido $|a, b|$, che può essere aperto, semi-aperto, o chiuso; in esso è incluso l'1-taglio $[r, s] = \{x : I(x) = 1\}$. La funzione di appartenenza $I(x)$ è non decrescente sull'intervallo $[a, r]$ (che può essere vuoto), ed è non crescente sull'intervallo $[s, b]$ (che può anch'esso essere vuoto). Gli intervalli chiusi nitidi $[r, s]$ si ritrovano quando il supporto si riduce al nucleo.

Esercizio. Se I e U sono due intervalli sfocati, per i loro α-tagli si ha $(I+U)_\alpha = I_\alpha + U_\alpha$ e $(IU)_\alpha = I_\alpha U_\alpha$. Queste proprietà valgono più in generale per quantità sfocate a supporto limitato, i cui α-tagli siano insiemi chiusi.

Definizione. Un *numero reale sfocato* è un intervallo sfocato la cui funzione caratteristica vale 1 in un unico numero reale nitido r (si continuano a usare scritture del tipo di $\tilde{r}$, circa r).

4 bis: Il modus ponens sfocato, ossia: la dinamica della logica sfocata.

Il **modus ponens** sfocato, pur ispirandosi all'omonimo sillogismo aristotelico, quello ben noto sulla mortalità di Socrate, se ne allontana parecchio, per cui è meglio introdurlo su un esempio che già nasca sfocato (*vago*, o anche *linguistico*, con riferimento al linguaggio ordinario, non simbolico, e alla logica del *buon senso*, o *common sense logic*, che del linguaggio ordinario sa servirsi con incredibili e quotidiani successi pratici):

Premessa: X è circa 3
Regola d'inferenza[1]: Se X è grande, allora Y è piccolo
Conclusione: Y gode dell'attributo ??.

Il problema è quello di definire in maniera adeguata l'attributo incognito che figura nella conclusione (beninteso si tratta di un attributo "linguistico di natura numerica", proprio come "circa 3" nella premessa e "grande" e "piccolo" nella regola d'inferenza), in maniera da "esplicitare" le affermazioni su Y contenute, in maniera solo implicita, nella premessa e nella regola d'inferenza. Nel sillogismo, X e Y sono due *variabili* di natura numerica[2] che noi trattiamo come *variabili linguistiche*, tipiche della teoria fuzzy, nel senso che ad esse attribuiamo *attributi linguistici* vaghi come: grande, molto grande, esageratamente grande, abbastanza diverso da 2, ecc. Questi attributi linguistici vengono descritti (codificati, *numerizzati*) tramite insiemi sfocati sull'universo nitido dei numeri reali. Nel caso di X abbiamo due attributi, $A(x)$ e $A'(x)$, che figurano rispettivamente nella premessa e nell'antecedente, mentre dei due attributi di Y, $B(y)$ nel conseguente e $B'(y)$ nella conclusione, il secondo va appena costruito. In astratto il sillogismo sfocato ha dunque il seguente formato:

Premessa: La variabile X gode dell'attributo $A(x)$

[1] Una tale regola ha la struttura: antecedente implica conseguente.
[2] Le loro specifiche concrete (le loro determinazioni possibili) sono numeri reali. Il termine variabile viene usata in maniera simile a quando si parla di variabili aleatorie, anch'esse affette da conoscenza incompleta. La variabile linguistica X e il punteggio che sta per uscire su un dado sono comunque "quantità" numeriche affette da conoscenza incompleta, vaghezza della descrizione che figura nella premessa, oppure incertezza di tipo aleatorio, probabilistico.

Regola d'inferenza: Se la variabile X gode dell'attributo $A'(x)$, allora la variabile Y gode dell'attributo $B(y)$

Conclusione: La variabile Y gode dell'attributo $B'(y) = \psi\big(A(x), A'(x), B(y)\big)$ (da calcolare).

La funzione ψ sta ad indicare che i gradi di appartenenza sfocati $B'(y)$ dei numeri nitidi y all'insieme sfocato B', ossia i valori logici sfocati[3] $B'(y)$, devono essere calcolabili[4] a partire dai gradi di appartenenza $A(x)$, $A'(x)$ e $B(y)$. Si tratta di specificare ψ in maniera sensata. Pensiamo per un momento a una situazione limite di tipo nitido, in cui i valori logici in gioco siano o 0 o 1, senza sfumature intermedie. Che cosa si puo' "nitidamente" dire della variabile Y? Costruiamo l'attributo $B'(y)$ a valori 0 e 1; quand'e' $B'(y) = 1$? Risposta: deve esistere una specifica x di X tale che la premessa sia vera in x, e tale che la y sia implicato dalla x. Più in esteso: OR x_1 è vero AND x_1 implica y, OR x_2 è vero AND x_2 implica y, OR $\ldots$, OR x_i è vero AND x_i implica y, OR $\ldots$. Abbiamo scritto i connettivi in inglese per rendere ancora più evidente la natura logica[5] di quello che stiamo facendo. In simboli,[6] e non si richiede più che i valori logici siano nitidi, la variabile Y gode dell'attributo:

$$B'(y) \;=\; \max_x \big[A(x) \wedge I(x,y)\big]$$

dove $I(x,y)$ è il valore logico della regola di inferenza calcolato in x e y: tale valore logico dipende *solo* dai valori logici $A'(x)$ dell'antecedente in x e $B(y)$ del conseguente in y: $I(x,y) = I\big(A'(x), B(y)\big)$. Se si opta per l'*implicazione di Mamdani*, si ha $I(x,y) = A'(x) \wedge B(y)$, e dunque

$$B'(y) \;=\; \max_x \big[A(x) \wedge A'(x) \wedge B(y)\big] \;=\; h \wedge B(y)$$

dove la *costante* h è definita da: $h = \max_x \big[A(x) \wedge A'(x)\big]$. In pratica l'attributo B' è identico all'attributo B, solo "tagliato" ad altezza h. Un'osservazione: a ritroso potremmo chiederci qual sia l'attributo linguistico della lingua italiana che viene "numerizzato" dall'insieme sfocato B' che abbiamo appena costruito. Ovviamente tale attributo non

---

[3] Come al solito, "confondiamo" insiemistica ("il grado di appartenenza di x all'insieme A è di 3/4") e logica ("il valore logico della proposizione x appartiene all'insieme A è di 3/4"), e quindi "confondiamo" le operazioni insiemistiche di unione, intersezione e complementazione con quelle logiche di disgiunzione OR, congiunzione AND e negazione NOT.

[4] Anzi: *facilmente* calcolabili, com'è tipico della *soft mathematics* (della *matematica flessibile*).

[5] Il lettore riconoscera un argomentazione simile a quella che porta a formulare il *principio di estensione*, che serve a trasformare funzioni fra quantità nitide in funzioni fra quantità sfocate.

[6] A rigore bisognerebbe scrivere l'estremo superiore sup invece del massimo max; in pratica nel controllo sfocato si usano quasi solo rozze funzioni di appartenenza trapezoidali che garantiscono a iosa l'esistenza del massimo, senza dover ricorrere al suo "surrogato tecnico".

è verosimile pescarlo nello Zingarelli, ma potremmo inventarlo, e decidere che il neologismo *grango* sia un aggettivo ben numerizzato da B', nello stesso senso in cui l'aggettivo grande è ben numerizzato da B. Allora il nostro ragionamento diventa, in buon toscano ampliato:

Se X è circa 3, e se X grande implica Y piccolo, allora Y è grango,

ciò che ne evidenzia la natura linguistica, da linguaggio ordinario e da logica del buon senso. Attenzione: ci siamo resi colpevoli di quattro violenze, che sono tre *numerizzazioni* (di "circa 3", di "grande" e di "piccolo"), e una *verbalizzazione* (che ci ha fatto inventare l'aggettivo "grango"). La quarta violenza viene di solito evitata, e rimane soltanto "implicita".

Esempio. Codifichiamo (numerizziamo) "circa tre" con il triangolo $(2, 3, 4)$, "grande" con il trapezio $(2, 5, +\infty, +\infty)$ e "piccolo" con il trapezio $(0, 0, 3, 5)$. "Grango" è come "piccolo", ma "tagliato" all'altezza $h = 1/2$.

5. Entropie sfocate

Torniamo a un universo del discorso finito, $\mathcal{X} = \{x_1, x_2, \ldots, x_K\}$. De Luca e Termini hanno proposto di misurare la *sfocatezza* di un insieme sfocato $A = (f_1, f_2, \ldots, f_K)$, definito dai gradi di apparteneza $f_i = A(x_i)$, mediante l'*entropia sfocata*

$$H_{\mathrm{Sh}}(A) = \sum_{1 \le i \le K} h(f_i)$$

L'uso dell'entropia di Shannon $h(f_i) = -\log_2 f_i - (1 - f_i)\log_2(1 - f_i)$ in un contesto il cui oggetto è la *vaghezza*, più che l'*incertezza* probabilistica, è opinabile. Ad ogni buon conto, l'entropia sfocata $H_{\mathrm{Sh}}(A)$ è compresa fra 0 e K; vale 0 sse A è nitido; vale K sse A è l'insieme "totalmente sfocato" con $f_i = \frac{1}{2} \ \forall i$. C'è inoltre invarianza rispetto alla complementazione: $H_{\mathrm{Sh}}(A) = H_{\mathrm{Sh}}(\overline{A})$ (un insieme è tanto sfocato quanto lo è il suo complementare); più in generale: se due insiemi sfocati A e B hanno la stessa *sagoma*, hanno anche la stessa entropia sfocata (la sagoma $\mathcal{S}(A)$ di un insieme sfocato A è data dall'intersezione $A \cap \overline{A}$, ed è sempre inclusa nell'insieme totalmente sfocato; la sagoma $\mathcal{S}(A)$ è vuota sse A è nitido). Se la sagoma di A è inclusa nella sagoma di B (nel senso dell'inclusione fra insiemi sfocati), allora $H_{\mathrm{Sh}}(A) \le H_{\mathrm{Sh}}(B)$.

Del resto le stesse proprietà che abbiamo elencato vengono sodddisfatte da funzionali molto più semplici, ad esempio dalla *misura di sfocatezza triangolare*:

$$H_{\Delta}(A) = 2 \sum_{1 \le i \le K} \min(f_i, 1 - f_i)$$

che è ben interpretabile, visto che $\min(f_i, 1-f_i)$ può essere visto come il valore logico della proposizione: "il grado di appartenenza di x_i ad A *non* è nitido" (□! Cenno: la proposizione "il grado di appartenenza di x_i ad A è nitido" si può rienunciare affermando "$x_i \in A$ oppure $x_i \notin A$", in cui compare l'operazione logica di disgiunzione e il cui valore logico di verità è

$\max[f_i, 1-f_i]$; negando si ottiene appunto $1 - \max[f_i, 1-f_i] = \min[f_i, 1-f_i]$). Il fattore 2 è stato inserito solo per scalare i valori di $H_\Delta(A)$ all'intervallo $[0, K]$).

Esercizio. La *numerosità* o la *cardinalità* $|A|$ di un insieme sfocato F definito tramite i gradi di appartenenza $(f_1, f_2, \ldots, f_K)$ è data dalla somma $\sum_i f_i$; il concetto estende quello nitido. Si ha: $H_\Delta(A) = 2|\mathcal{S}(A)|$.

Siamo ormai in grado di distinguere fra tre concetti pertinenti nella descrizione di situazioni di *conoscenza parziale* (*partial knowledge*): l'*incertezza*, l'*ignoranza* (che alcuni chiamano pudicamente la *non-specificità*) e la *vaghezza*.
L'incertezza compete al calcolo delle probabilità nella sua variante soggettiva o "epistemica"; l'incertezza massima è associata alle probabilità uniformi.
L'ignoranza, o la non-specificità, competono al calcolo delle possibilità (e anche alla teoria degli insiemi sfocati); l'ignoranza massima è associata alle possibilità uniformemente uguali a 1.
La vaghezza compete alla teoria degli insiemi sfocati; la vaghezza massima è associata agli insiemi sfocati la cui funzione caratteristica è costante e uguale a 1/2.

6. Teoria dell'attestabilità (evidence theory)

Sul piano epistemico (*episteme* = conoscenza), la probabilità dell'evento $\{X \in A\}$, $A \subseteq \mathcal{X}$, viene interpretata come il *grado di fiducia* che l'agente umano ha nella realizzazione di A.

> In questo contesto l'agente umano viene di solito chiamato l'*esperto*, e non lo *sperimentatore* come nelle probabilità "oggettive" standard.

Si è sostenuto che il calcolo delle probabilità standard sarebbe insufficiente per dar ragione di questa interpretazione epistemica della probabilità. Si è proposto di ampliare il calcolo standard e di assegnare un vettore di probabilità m (*distribuzione delle attestazioni* o *basic probability assignment* in inglese) su $2^{\mathcal{X}} - \emptyset$ invece che su $\mathcal{X}$. Il significato del numero $m(A)$ è il *peso delle attestazioni* che l'esperto ha in favore di A, mentre non sa come ripartire questo favore fra i sottinsiemi stretti di A. La distribuzione delle attestazioni m definisce un *corpus di attestazioni*, o BOE (*body of evidence*), su $\mathcal{X}$; gli eventi o sottinsiemi F per cui $m(F)$ è strettamente positivo sono chiamati *focali*. Si noti che formalmente m è un vettore di probabilità sull'insieme dei sottinsiemi di $\mathcal{X}$, ma concettualmente non è affatto una probabilità. Altri due numeri vengono associati ad A, la sua *credibilità* (*belief*) e la sua *plausibilità* (*plausibility*):

$$\mathrm{Bel}(A) = \textstyle\sum_{B \subseteq A} m(B) \quad , \quad \mathrm{Pl}(A) = \textstyle\sum_{B \cap A \neq \emptyset} m(B)$$

Non è m, ma sono questi due numeri, casomai, che fanno le veci della probabilità di A, $\mathrm{Prob}(A)$: in effetti, alcuni vedono l'intervallo $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$ come la *probabilità intervallare* di A, laddove la *probabilità puntuale* rimane non specificata. Nel calcolo della credibilità di A intervengono i focali che lo convalidano direttamente (lo implicano), mentre nel calcolo della sua plausibilità intervengono i focali che non lo contraddicono (non lo negano); la credibilità di A è il peso globale delle attestazioni a suo favore, la plausibilità di A è il peso globale delle attestazioni che non gli sono contrarie.

Nel caso speciale in cui m assegni pesi positivi solo ai singoletti, credibilità e plausibilità sono uguali, e la probabilità intervallare di A si riduce a una probabilità puntuale. Dunque le probabilità epistemiche standard (le probabilità *bayesiane*) si ritrovano in questo caso speciale. Se $m(\mathcal{X}) = 1$ il BOE unifocale è *vacuo* e descrive una situazione di *ignoranza totale*: ogni sottinsieme proprio di $\mathcal{X}$ ha credibilità 0 e plausibilità 1. Più in generale, un BOE si chiama *unifocale* quando m concentra l'intero peso unitario su un'unico evento F.

Esempi numerici. Sull'universo $\{a, b, c, d\}$ sia dato il BOE $m(a, b) = 2/3$, $m(c, d) = 1/3$; l'assegnazione è di fatto una probabilità bayesiana sull'universo "ingrossato" (visto rozzamente) i cui *due* soli elementi sono $\{a, b\}$ e $\{c, d\}$. Si ha $\mathrm{Bel}(a) = 0 < \mathrm{Pl}(a) = 2/3$, $\mathrm{Bel}(a, b) = \mathrm{Pl}(a, b) = 2/3$. Sia invece $\mu(a) = 2/3$, $\mu(a, b, c) = 1/3$: l'assegnazione non ha più nulla di bayesiano; stavolta è $\mathrm{Bel}(a) = 2/3 < \mathrm{Pl}(a) = 1$.

Si noti che la formula di Möbius (nella sua "variante" insiemistica; cfr. l'appendice) dà : $m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mathrm{Bel}(B)$, ciò che permette di ricostruire i pesi m a partire dalle credibilità .

Altre formule notevoli ($\square$!) sono:

$\mathrm{Bel}(A) + \mathrm{Pl}(\overline{A}) = 1$ (dualità)
$A \cap B = \emptyset \Rightarrow \mathrm{Bel}(A \cup B) \geq \mathrm{Bel}(A) + \mathrm{Bel}(B)$ (superadditività della credibilità)
$A \cap B = \emptyset \Rightarrow \mathrm{Pl}(A \cup B) \leq \mathrm{Pl}(A) + \mathrm{Pl}(B)$ (subadditività della plausibilità)

In particolare, può essere $\mathrm{Bel}(A) + \mathrm{Bel}(\overline{A}) < 1$, ed è a questa mancata uguaglianza che ci si riferisce quando si dice che le credibilità sono "probabilità non additive".

> Ci sono precedenti storici di probabilità non additive - e dunque eretiche per i probabilisti bachettoni - già nell'*Ars Conjectandi* di Bernoulli e nel *Neues Organon* di Lambert del 1764. Usare il termine probabilità per funzionali che non sono additivi crea confusione; al di là delle parole, tuttavia, la natura epistemica e "probabilistica" delle credibilità e delle plausibilità è innegabile. Etimologicamente "probabilile" e "attestabile" sono sinonimi.

Esempio: Il vaso Ming.

L'esempio che segue è dovuto a Glenn Shafer, uno dei padri, con Arthur Dempster, della teoria non-additiva dell'attestabilità. Si deve stabilire se un bel vaso cinese sia un vaso Ming autentico, o un vaso Ming falso. Quanto vale Prob{il vaso è autentico}? Introduciamo due personaggi, Tizio, grande esperto del periodo Ming, e Caio, la cui cultura orientale si limita a qualche scontata banalità. Supponiamo che entrambi siano probabilisti tradizionali e che emettano lo stesso verdetto: Prob{il vaso è autentico} = Prob{il vaso non è autentico} = $\frac{1}{2}$. Gli stati di conoscenze sono tuttavia ben diversi: Caio semplicemente non sa che pesci pigliare, Tizio si è trovato invece a dover soppesare con la sua raffinata cultura indicazioni estremamente contradditorie: la qualità dello smalto lo faceva propendere per il sı, la forma dell'ansa per il no, e così via. Dunque non c'è modo, nell'impostazione tradizionale, di distinguere fra totale ignoranza e totale incertezza: entrambe conducono alla stessa valutazione numerica. Se Tizio e Caio fossero invece dempsteriani-shaferiani, avrebbero modo di pesare in maniera diversa le attestazioni che ciascuno di loro ha a favore o contro l'autenticità del vaso. Il verdetto di Tizio potrebbe rimanere lo stesso, ma quello di Caio esprimerebbe ora in maniera netta la totale mancanza di attestazioni concrete sia nell'uno che nell'altro senso: Bel{il vaso è autentico} = Bel{il vaso non è autentico} = 0. La totale ignoranza viene dunque codificata mediante due "probabilità " non-additive rigorosamente nulle, che sommano a zero. Attenzione: sul piano *operativo* delle scommesse tanto care ai neo-bayesiani (sul piano *decisionale*), i due stati di conoscenza, pur distinti, potrebbero ben portare Tizio e Caio alla stessa "azione", vale dire a una scommessa 1 contro 1. Talvolta si fa confusione fra due livelli, quello epistemico (statico) e quello operativo (dinamico) che ad esso consegue.

> Il vaso Ming è l'esempio emblematico di una situazione in cui il probabilista frequentista ("oggettivo") si rifiuterebbe di dare valutazione numeriche.

Un caso particolare notevole è quello delle *attestazioni consonanti*, quando gli insiemi focali sono *annidati* l'uno nell'altro, ossia formano una *catena*: $F_1 \subset F_2 \ldots \subset F_r$:

Teorema. La credibilità è una necessità se e solo se il BOE è consonante. Ossia, per dualità: la plausibilità è una possibilità se e solo se il BOE è consonante.

Dimostrazione. Dimostreremo quest'ultima affermazione.
Dato un BOE consonante, possiamo sempre supporre, riordinando al caso l'insieme universo $\mathcal{X}$, che i suoi focali siano *insiemi iniziali* del tipo $\{x_1, x_2, ..., x_i\}$ (attenzione: non tutti i K insiemi iniziali sono necessariamente anche insiemi focali, per qualche i potrebbe ben essere $m(\{x_1, x_2, ..., x_i\}) = 0$). Ma allora ($\square$!) le plausibilità $\mathrm{Pl}(x_i)$ dei K singoletti x_i sono ordinate in senso non crescente e $\mathrm{Pl}(x_1) = 1$ (x_1 sta in tutti i focali), per cui possiamo definire una distribuzione di possibilità ponendo $\pi_i = \mathrm{Pl}(x_i)$, $\pi_i \geq \pi_{i+1}$. Come subito si verifica: $\pi_1 = 1, \pi_2 = 1 - m(x_1, \ldots), \pi_3 = 1 - m(x_1, \ldots) - m(x_2, \ldots), \pi_4 = 1 - m(x_2, \ldots) - m(x_2, \ldots) - m(x_3, \ldots)$. Sia A un evento e sia x_j il suo primo elemento, cioè quello di indice più basso. I focali F sono insiemi iniziali, e dunque $x_j \in F$ se e solo se $A \cap F \neq \emptyset$, per cui

$$\mathrm{Pl}(A) = \textstyle\sum_{F: A \cap F \neq \emptyset} m(F) = \sum_{F: x_j \in F} m(F) = \mathrm{Pl}(x_j) \equiv \pi_j = \max_{x_i \in A} \pi_i = \Pi(A)$$

In particolrae: . Viceversa, dato un vettore di possibilità, senza effettiva restrizione possiamo assumere che esso sia ordinato in senso non decrescente: $1 = \pi_1 \geq \pi_2 \geq \ldots \geq \pi_K$. Costruiamo un BOE consonante i cui focali sono segmenti iniziali ponendo $m\{x_1, x_2, ..., x_i\} = \pi_i - \pi_{i+1}$, ($\pi_{K+1} \doteq 0$; m individua effettivamente un BOE perchè $\sum_F m(F) = \pi_1 = 1$). Si ha:

$$\mathrm{Pl}(A) = \textstyle\sum_{i \geq j} m(x_1, \ldots, x_i) = \sum_{i \geq j} (\pi_i - \pi_{i+1}) = \pi_j = \max_{x_i \in A} \pi_i = \Pi(A)$$

$\square$

Un altro caso notevole è quello delle *probabilità incomplete*, che sono in un certo qual modo una nozione di confine fra "probabilità standard" e "probabilità non standard" (non additive). In un *vettore di probabilità incompleto* P le K componenti non-negative p_i sono vincolate ad avere una somma *minore* o uguale a 1, e non proprio uguale a 1. Posto $\sigma(A) = \sum_A p_i$, $1 - \sigma(\mathcal{X})$ è detto il *deficit* del vettore incompleto P. Nella teoria standard delle probabilità le probabilità incomplete sono giustificate con il concetto (empirico) della *non osservabilità* di certi esperimenti (ma nei casi peggiori sono poco più di un espediente matematico), mentre a esse si può dare un significato preciso all'interno della teoria dell'attestabilità. Diremo che un BOE è una probabilità incompleta quando gli unici focali possibili ($m(F) \neq 0$) sono i singoletti e l'insieme pieno $\mathcal{X}$; $m(\mathcal{X})$ è il deficit della probabilità incompleta, mentre $p_i = \mathrm{Bel}(x_i) = m(x_i)$. Più in generale, $\mathrm{Bel}(A) = \sigma(A)$ per $A \neq \mathcal{X}$, e $Pl(A) = \sigma(A) + m(\mathcal{X})$ per $A \neq \emptyset$ ($\square$!). Un deficit uguale a 1 dà il BOE vacuo, mentre deficit nulli danno le probabilità bayesiane (le probabilità complete). Le probabilità incomplete sono ben interpretabili empiricamente in termini di *non osservabilità*, quando di n esperimanti una percentuale pari a $m(\mathcal{X})$ non sia stata osservata. Allora, se A è un evento né pieno né vuoto, tutto ciò che si puo dire della sua probabilità empirica (frequenza di successo) è che essa è compresa fra $\mathrm{Bel}(A)$ e $\mathrm{Pl}(A)$.

Appendice: La formula di Möbius

La formula di inversione di Möbius viene di solito enunciata nella sua variante aritmetica sul reticolo degli interi ordinati per divisibilità. Noi la daremo nella variante insiemistica sul reticolo dei sottinsiemi ordinati per inclusione. Per evitare interruzioni banali osserviamo subito che $A \cap D = \emptyset$ implica $(-1)^{|A\cup D|} = (-1)^{|A|}(-1)^{|D|}$, e $B \subseteq A$ implica $(-1)^{|A-B|} = (-1)^{|A|}(-1)^{|B|}$ ($\square$!).

Lemma. $\sum_{A:A\subseteq C}(-1)^{|A|} = 1$ se e solo se $C = \emptyset$, altrimenti vale 0.
Dimostrazione. Per $|C| \geq 1$ sviluppa il binomio $(1-1)^{|C|} = 0$.

Lemma. Sia $A \subseteq C$. Allora $\sum_{B:A\subseteq B\subseteq C}(-1)^{|B|} = (-1)^{|A|}$ se e solo se $A = C$, altrimenti vale 0.
Dimostrazione. Segue dal precedente, poichè :

$$\sum_{B:A\subseteq B\subseteq C}(-1)^{|B|} = \sum_{D:D\subseteq C-A}(-1)^{|A\cup D|} = (-1)^A \sum_{D:D\subseteq C-A}(-1)^D$$

Formula di Möbius: Sia $g : 2^{\mathcal{X}} \to \Re$, e sia $f(A) = \sum_{B\subseteq A} g(B)$. Allora:
$g(A) = \sum_{B:B\subseteq A}(-1)^{|A-B|}f(B)$.
Dimostrazione. Stante la definizione di $f(B)$, e grazie all'ultimo lemma:

$$\sum_{B:B\subseteq A}(-1)^{|A-B|}f(B) =$$
$$= (-1)^{|A|}\sum_{B:B\subseteq A}(-1)^{|B|}f(B)$$
$$= (-1)^{|A|}\sum_{B:B\subseteq A}(-1)^{|B|}\sum_{C:C\subseteq B} g(C)$$
$$= (-1)^{|A|}\sum_{C:C\subseteq A} g(C)\sum_{B:C\subseteq B\subseteq A}(-1)^{|B|} = (-1)^{|A|}g(A)(-1)^{|A|} = g(A).$$

7. La regola di Dempster e gli universi aperti

Un problema cruciale nella teoria dell'attestabilità è quello di *sintetizzare* le attestazioni più o meno conflittuali di due esperti distinti in un'unica "superattestazione". Non ci si può accontentare di una *statica* che descrive stati di conoscenza, bisogna passare a una *dinamica* che sapppia aggiornarli e gestirli.

Partiamo da un esempio reale, anche se semplificato. Si tratta della diagnosi di un caso di itterizia colestatica. L'itterizia (mal giallo) è dovuta a un pigmento prodotto dal fegato e trasportato nel sangue, la bilirubina. L'eccesso di bilirubina può esser dovuto o a una malattia all'interno del fegato (colestasi intraepatica), o al blocco dei dotti biliari (colestasi extraepatica). Casi possibili della colestasi intraepatica sono l'epatite Ep e la cirrosi Ci ; casi possibili della colestasi extraepatica sono i calcoli Cl e il cancro pancreatico Pa (non ne considereremo altri). Ci sono dunque 4 singoletti: $\mathcal{X}$ = itterizia = {Ep, Ci, Cl, Pa}; ci sono inoltre due eventi (sottinsiemi) notevoli: colestasi intraepatica = {Ep, Ci} e colestasi extraepatica = {Cl, Pa}.

Prima attestazione (prima diagnosi): Dopo aver visitato il paziente, il primo medico attesta quanto segue: ci sono forti indicazioni che il malato soffra di colestasi intraepatica, *di conseguenza* è inverosimile che il malato soffra di colestasi extraepatica. Codifica numerica (stiamo di nuovo dando i voti come si fa a scuola, ma non è il medico che "dà i numeri", la codifica numerica della sua diagnosi verbale è compito nostro): $m_1\{\mathrm{Ep}, \mathrm{Ci}\} = .8$, $m_1\{\mathrm{Cl}, \mathrm{Pa}\} = .2$. Commento: l'attestazione è "pressocchè " bayesiana, solo su un alfabeto "ingrossato" di due elementi (i due focali, colestasi intraepatica e colestasi extraepatica, sono disgiunti ed esauriscono lo spazio). Per esempio: Bel_1(cirrosi) = 0, Pl_1(cirrosi) = .8, Bel_1(colestasi intraepatica) = Pl_1(colestasi intraepatica) = .8. La situazione essendo "quasi" bayesiana, nulla osta a vedere il numero .8 come il valore di una scommessa in cui si punta 8 contro 2.

Seconda attestazione (seconda diagnosi). Il secondo medico attesta quanto segue: il malato non soffre di epatite, anche se ciò non si può del tutto escludere; *non si impegna* sul fatto se questa indicazione influenzi o no la credibilità delle altre ipotesi in gioco. Codifica numerica: $m_2\{\mathrm{Ci}, \mathrm{Cl}, \mathrm{Pa}\} = .9$, $m_2\{\mathcal{X}\} = .1$. Commento: l'attestazione è irriducibilmente non-bayesiana; in un quadro bayesiano avremmo dovuto assegnare il peso mancante .1 a Ep, e non evasivamente a $\mathcal{X}$, come abbiamo fatto. Rifletti al "di conseguenza" che compare nella prima attestazione e al non-impegno della seconda. Si ha, per esempio: Bel_2(cirrosi) = 0, Pl_2(cirrosi) = 1, Bel_2(colestasi intraepatica) = 0, Pl_2(colestasi intraepatica) = 1. Per inciso: i focali sono annidati e dunque, almeno formalmente, l'attestazione è di tipo possibilistico.

Se la *rappresentazione* della conoscenza è soddisfaciente, la sua *gestione*, come ora vedremo, è problematica (la teoria ha una buona "statica", ma la "dinamica" è opinabile).

Sintesi delle attestazioni (*evidence pooling*). La *regola di Dempster* permette di sintetizzare, o conglobare, in un'unica attestazione due attestazioni *indipendenti* e di *uguale importanza*;

essa è ispirata alla regola del prodotto per le probabilità di eventi indipendenti. Per il momento ci limitiamo al caso di due attestazioni *non dissonanti*, come quelle dell'esempio, quando cioè ciascun focale della prima attestazione interseca ciascun focale della seconda. Si pone allora, per ogni evento $A \neq \emptyset$ di $\mathcal{X}$:

$$m(A) \equiv m_{1\otimes 2}(A) = \sum_{B,C:\ B\cap C=A} m_1(B) \times m_2(C)$$

Nel nostro caso si trova: $m\{\mathrm{Ci}\} = .72$, $m\{\mathrm{Ep,Ci}\} = .08$, $m\{\mathrm{Ca,Pa}\} = .18 + .02 = .2$. Si ha: $\mathrm{Bel}_{1\otimes 2}(\text{cirrosi}) = .72$, $\mathrm{Pl}_{1\otimes 2}(\text{cirrosi}) = .8$, $\mathrm{Bel}_{1\otimes 2}(\text{colestasi intraepatica}) = .8$, $\mathrm{Pl}_{1\otimes 2}(\text{colestasi intraepatica}) = .8$.

Se le due attestazioni sono dissonanti per arrivare a 1 manca il peso $m(\emptyset) = \sum_{B\cap C=\emptyset} m_1(B) \times m_2(C)$. Che fare? La soluzione proposta da Dempster è molto semplicemente quella di aggiungere nella formula un coefficiente di normalizzazione, e cioè di porre al secondo membro il denominatore $1 - m(\emptyset)$. Si tratta di un mero espediente, che è stato assai contestato, e che può portare a risultati contro-intuitivi, come mostra l'esempietto seguente, in cui le due attestazioni sono "fortemente dissonanti" o "contraddittorie". Una via d'uscita molto più radicale e interessante è quella del *mondo aperto* (*cornice aperta*, *open frame*), che dà la possibilità di assegnare un peso m positivo anche all'evento vuoto $\emptyset$. Se si accetta questo ampliamento, la regola viene estesa com'è stata scritta anche a $A = \emptyset$: $m(\emptyset) = m_\emptyset$. Porre $m(\emptyset) > 0$ equivale ad affermare che la "cornice di riferimento" $\mathcal{X}$ è incompleta; ciò (ma stiamo entrando in un terreno minato!) segnala la presenza di uno stato di conoscenze *contraddittorio*.

Esempietto giudiziario. $\mathcal{X} = \{\text{Zoe, Zamfira, Zuleica}\}$: una delle tre è colpevole di aver ammazzato la vittima. Il testimone attendibile A scagiona Zuleica, e attesta $m_A(\text{Zoe}) = 0.99$, $m_A(\text{Zamfira}) = 0.01$; il testimone altrettanto attendibile B scagiona Zoe, e attesta invece $m_B(\text{Zuleica}) = 0.99$, $m_B(\text{Zamfira}) = 0.01$. Se si adopera il coefficiente di normalizzazione, la dissonanza fra le due testimonianze non impedisce di far dichiarare Zamfira "certamente" colpevole, degna di venir condannata alla massima pena: $m_{A\otimes B}(\text{Zamfira}) = 1$!

> Con forza polemica i sostenitori dell'attestabilità parlano di BOE *non dogmatici*, quando la cornice di riferimento ha essa stessa massa positiva (come nella seconda diagnosi). Il passaggio dai "mondi" chiusi a quelli aperti è, se si vuole, un ulteriore passo nella "sdogmatizzazione" delle teorie probabilistiche "ortodosse".

Nel seguito considereremo solo universi chiusi: $m(\emptyset) = 0$.

8. Attestazioni e pseudo-entropie

Rammentiamo uno scenario standard nelle probabilità oggettive. Un esperimento aleatorio X con K risultati possibili viene eseguito ed è retto dal vettore di probabilità $P = (p_1, p_2, \ldots, p_K)$; $X \in \mathcal{X} = (x_1, x_2, \ldots, x_K)$. Come sappiamo, l'*entropia di Shannon* della variabile aleatoria X (equivalentemente, l'entropia di Shannon del vettore di probabilità P), definita dalla:

$$H(X) = H(P) = -\sum_{1\leq i\leq K} p_i \log p_i$$

viene interpretata come una misura adeguata dell'incertezza che lo sperimentatore ha prima di eseguire l'esperimento.

Oltre all'entropia di Shannon, c'è un'altra misura d'informazione "storica", la misura logaritmica di *Hartley* del 1928:

$$\mathrm{Hartley}(\mathcal{X}) = \log|\mathcal{X}|\,, \quad \mathcal{X} \neq \emptyset$$

legata alla numerosità, o, come preferiremo d'ora in poi dire, alla *non-specificità* dell'insieme $\mathcal{X}$ (su cui non occorre sia assegnata una struttura probabilistica).

Parecchie misure d'informazione sono state proposte per i BOE, come, per esempio, le misure di non-specificità, di *dissonanza* e di *confusione* (o *conflittualità*) che seguono (*pseudo-entropie*). Esse sono palesemente ispirate, per analogia, all'entropia di Shannon e alla misura di Hartley. Sotto m è la distribuzione delle attestazioni che definiscono il BOE; abbiamo usato la lettera F in indice per sottolineare che le somme sono di fatto limitate ai focali:

$N(m) = \sum_F m(F) \log|F|$: misura la non-specificità (l'ignoranza)
$D(m) = -\sum_F m(F) \log \mathrm{Pl}(F)$: misura la dissonanza
$C(m) = -\sum_F m(F) \log \mathrm{Bel}(F)$: misura la confusione (la conflittualità)

Esempietti numerici. Riprendiamo sull'universo $\{a, b, c, d\}$ le due attestazioni $m(a, b) = 2/3$, $m(c, d) = 1/3$ e $\mu(a) = 2/3$, $\mu(a, b, c) = 1/3$. Si ha $N(m) = 2/3$, $D(m) = C(m) = \log 3 - 2/3$, $N(\mu) = 1/3 \log 3$, $D(\mu) = 0$, $C(\mu) = 2/3 \log 3 - 2/3$, tutto in bit.

Esempio. Prendiamo una probabilità incompleta $(p_i, p_2, \ldots, p_K)$ di deficit $m(\mathcal{X})$. Si ha $N(m) = m(\mathcal{X}) \log|\mathcal{X}|$, $D(m) = -\sum_i p_i \log\big(p_i + m(\mathcal{X})\big)$, $C(m) = -\sum_i p_i \log p_i$.

Purtroppo il carattere piuttosto ad hoc delle ultime due proposte, $D(m)$ e $C(m)$, è ormai ammesso, ciò che ne rende controverso l'uso. Sotto tiro è specialmente la misura di confusione $C(m)$; del resto la distinzione fra dissonanza e confusione è essa stessa sfuggente. Le due misure "storiche", di Shannon e di Hartley, si ritrovano nei casi limite dei BOE bayesiani e del BOE vacuo (totale ignoranza: tutto quel che si sa è il fatto scontato che l'universo è $\mathcal{X}$, e basta).

Caso bayesiano: $N(m) = 0$, $D(m) = C(m) = H(P)$
Totale ignoranza: $N(m) = \mathrm{Hartley}(\mathcal{X}) = \log|\mathcal{X}|$, $D(m) = C(m) = 0$

Per valutare l'adeguatezza delle misure d'informazione, è essenziale esaminare il loro campo di variazione e capire il significato delle condizioni di uguaglianza che garantiscono il raggiungimento dei valori estremi. Elenchiamo le disuguaglianze fondamentali:

$$0 \leq N(m) \leq \log |\mathcal{X}|$$
$$0 \leq D(m) \leq \log |\mathcal{X}|$$
$$0 \leq C(m) < |\mathcal{X}|$$

Purtroppo l'ultima disuguaglianza è stretta e la lasciamo in sospeso; ad essa torneremo fra poco per i commenti e la dimostrazione. Seguono i criteri di uguaglianza:

$N(m) = 0$ se e solo se la m è bayesiana
$D(m) = 0$ se e solo se non ci sono focali disgiunti (per esempio nel caso consonante)
$C(m) = 0$ se e solo se la m è unifocale

$N(m) = \log |\mathcal{X}|$ se e solo se la m è vacua (caso della totale ignoranza)
$D(m) = \log |\mathcal{X}|$ se e solo se la m è bayesiana uniforme

Dimostrazioni. Sono immediate ($\square$!); vedremo solo la $D(m) \leq \log |\mathcal{X}|$. Introduciamo un vettore di probabilità P su $\mathcal{X}$ "frazionando uniformemente sui singoletti la massa di ciascun focale", ossia ponendo:

$$P(x_i) = \sum_{F: x_i \in F} \frac{m(F)}{|F|}$$

Le $P(x_i)$ sommano a 1 ($\square$!). Ovviamente per $x_i \in F$, $\mathrm{Pl}(F) \geq \mathrm{Pl}(x_i) \geq P(x_i)$, per cui: $D(m) = -\sum_{x_i} \sum_{F: x_i \in F} \frac{m(F)}{|F|} \log \mathrm{Pl}(F) \leq -\sum_{x_i} \log P(x_i) \sum_{F: x_i \in F} \frac{m(F)}{|F|} = H(P) \leq \log |\mathcal{X}|$.
Per avere l'uguaglianza $D(m) = H(P)$ dev'essere $\mathrm{Pl}(x_i) = P(x_i)$, ossia $|F| = 1$ per tutti i focali F, e dunque il BOE m dev'essere bayesiano; per avere l'uguaglianza $D(m) = H(P) = \log |\mathcal{X}|$ il BOE bayesiano dev'essere uniforme. $\square$

La pseudo-entropia $N(m)$ non è sempre maggiore o sempre minore nè della $D(m)$ nè della $C(m)$ (basta pensare ai BOE bayesiani e al BOE vacuo). Invece ($\square$!):

$D(m) \leq C(m)$
$D(m) = C(m)$ se e solo se i focali sono tutti disgiunti

I risultati trovati sono soddisfacienti dal punto di vista "filosofico", e convalidano, anche se non in maniera cogente, le nostre interpretazioni in termini di misure della non-specificità, della dissonanza e della conflittualità di un corpus di attestazioni. Purtroppo però non è chiaro quale sia il massimo della misura di confusione, $C(m)$. La $C(m) \leq K$, anzi la $C(m) < K$, è subito implicata dalla

$$C(m) \leq H(m) = -\sum_{F} m(F) \log m(F) \leq \log(2^K - 1) < K\,,$$

dove $H(m)$ è l'entropia di Shannon della m vista come un "vettore di probabilità" a $2^K - 1$ componenti. Si ha l'uguaglianza $C(m) = H(m)$ se e solo se i focali formano un'anticatena

($\Box$!). Perfino la disuguaglianza $\forall m \; C(m) < \log(2^K - 1)$ è stretta, perchè l'entropia di Shannon $H(m)$ è massima in corrispondenza a un BOE, quello "uniforme" su $2^{\mathcal{X}} - \emptyset$, che *non* forma un'anticatena. In base al lemma di Sperner esposto in Appendice le anticatene massimali hanno cardinalità $\binom{K}{\lfloor \frac{K}{2} \rfloor}$; se distribuiamo il peso unitario m in maniera uniforme su tali anticatene otteniamo:

$$\log \binom{K}{\lfloor \frac{K}{2} \rfloor} \leq \max_m C(m) < \log(2^K - 1)$$

Neppure la limitazione inferiore non è il vero massimo, come mostreremo fra poco studiando il caso delle probabilità incomplete: il problema del massimo di $C(m)$ rimane insomma e purtroppo aperto. Notiamo tuttavia che quanto sappiamo basta per dimostrare che

$$\max C(m) \approx K \quad \text{per} \quad K \gg 1$$

Per provarlo basta usare l'approssimazione asintotica del coefficiente binomiale basata sull'entropia:

$$\binom{K}{r} = 2^{K\{h(\frac{r}{K}) - \epsilon_K\}}$$

dove $h(x) = H(x, 1-x)$, $\epsilon_K \to 0$.

Se si confrontano dissonanza e conflittualità, si nota che $\max C(m) \approx K \gg \max D(m) = \log K$.

Che la limitazione inferiore basata sul lemma di Sperner fosse il vero massimo della misura di confusione è stato a suo tempo congetturato; peccato non lo sia, sarebbe stato un massimo significativo!

Complementi. Nel caso delle probabilità incomplete, la misura di confusione $C(m)$ assume la forma:

$$C(m) = C(P) = -\sum_i p_i \log p_i$$

Nella letteratura $C(P)$ viene chiamata *entropia incompleta*. L'entropia incompleta $C(P)$ è non-negativa ed è nulla quando P è vacuo, o quando P è completo e deterministico ($\Box$!). Soffermiamoci sulla limitazione superiore. Cominciamo osservando che un vettore di probabilità incompleto ma non-vacuo P può venir reso completo dividendo le sue componenti per $\sigma = \sigma(\mathcal{X}) = 1 - m(\mathcal{X})$. Ciò fatto, l'ovvia limitazione superiore per l'entropia di Shannon completa, $H(P) \leq \log K$, dà per le entropie incomplete:

$$C(P) \leq -\sigma \log \frac{\sigma}{K}$$

con l'uguale se e solo se il vettore P è costante. Se si massimizza su σ nell'intervallo $]0, 1]$, si ritrova la vecchia limitazione per l'entropia completa nel caso $K \geq 3$: $C(P) \leq \log K$ con l'uguale se e solo se P è completo e uniforme. Per $K = 2$ c'è invece una sorpresa. Si scopre che $-\sigma \log \frac{\sigma}{K}$ supera $\log 2 = 1$ nel subintervallo $\frac{1}{2} < \sigma < 1$, e assume il valore massimo per

$P = (e^{-1}, e^{-1})$, quando $H(P) = \frac{2}{e \ln 2} \approx 1.06$. In particolare, in questo subintervallo $C(P)$ batte $\log \binom{2}{1} = 1$ tutte le volte che il vettore incompleto P è costante.
Per mancanza di focali, nel caso $K = 2$ tutti i BOE m sono probabilità incomplete, per cui studiare $C(P)$ equivale a studiare $C(m)$. I massimi e i minimi delle misure d'informazione dovrebbero avere un significato chiaramente interpretabile; invece il significato della limitazione superiore per $C(m)$ nel caso $K = 2$ è decisamente oscuro. Questo fatto getta gravi dubbi sull'adeguatezza della misura d'informazione pseudo-entropica $C(m)$.

Appendice: Il lemma di Sperner

Un'anticatena è una famiglia $\mathcal{F}$ di sottinsiemi $F \subseteq \mathcal{X}$ fra i quali è proibita la relazione di inclusione. *Esempio*: fissa r $(0 \le r \le K)$ e considera tutti i sottinsiemi di cardinalità r. L'esempio mostra che esistono anticatene di cardinalità $\binom{K}{r}$; in particolare, ricordando che il coefficiente binomiale più alto è quello centrale $(r = \frac{K}{2})$ se K è pari, i due centrali $(r = \frac{K \mp 1}{2})$ se K è dispari, abbiamo esibito anticatene di cardinalità, o numerosità, $\binom{K}{\lfloor \frac{K}{2} \rfloor}$. Il lemma di Sperner prova che non si può far di più.

Lemma di Sperner: Se $\mathcal{F}$ è un'anticatena, $|\mathcal{F}| \le \binom{K}{\lfloor \frac{K}{2} \rfloor}$.
Dimostrazione. Non daremo la dimostrazione filologica di Sperner del 1928, bensì quella di Lubell del 1966. Ci sono $K!$ permutazioni di $\mathcal{X}$; di queste $|F|!(K-|F|)!$ *cominciano* con F, nel senso che i primi $|F|$ elementi nella permutazione sono proprio quelli di F. Attenzione: la stessa permutazione non può cominciare con due insiemi distinti dell'anticatena $\mathcal{F}$! Dunque:

$$\sum_{F \in \mathcal{F}} |F|!(K - |F|)! \le K!$$

Indichiamo con n_r il numero di insiemi di $\mathcal{F}$ di cardinalità esattamente r. La relazione precedente si riscrive come:

$$\sum_r n_r r!(K - r)! \le K!$$

ossia

$$\sum_r \frac{n_r}{\binom{K}{r}} \le 1$$

Ma allora:

$$|\mathcal{F}| = \sum_r n_r \le \binom{K}{\lfloor \frac{K}{2} \rfloor} \sum_r \frac{n_r}{\binom{K}{r}} \le \binom{K}{\lfloor \frac{K}{2} \rfloor}$$

□

9. Probabilità imprecise e misure d'incertezza intervallari

Dal punto di vista tecnico, le *probabilità intervallari* fanno parte della *teoria delle capacità* proposta da Gustave Choquet negli anni 50 e mirata ad appplicazioni alla fisica, per inciso su "universi" di tipo continuo. Le probabilità intervallari costituiscono un quadro di riferimento molto ampio per la gestione dell'incertezza, forse troppo ampio per essere pratico.

E' dato un insieme chiuso $\mathcal{P}$ di vettori di probabilità P su $\mathcal{X}$, $\mathcal{P} \neq \emptyset$. Si pone rispettivamente

$$P_*(A) = \min_{P\in\mathcal{P}} P(A),$$
$$P^*(A) = \max_{P\in\mathcal{P}} P(A)$$

per definire la *probabilità inferiore* P_* e la *probabilità superiore* P^* di A. (Richiedere la chiusura di $\mathcal{P}$ serve essenzialmente a evitare il fastidio di scrivere *inf* e *sup* invece di *min* e *max*; oltre che chiuso $\mathcal{P}$ è anche compatto, visto che *qualunque* insieme di vettori di probabilità è limitato). Nella letteratura le *probabilità intervallari* vengono anche chiamate *probabilità imprecise*.

Le probabilità intervallari ammettono tre interpretazioni filosofiche, le prime due tutto sommato "ortodosse". Possono essere viste come *approssimazioni* o *stime* intervallari di probabilità puntuali mal conosciute. Si può assumere il punto di vista delle *opinioni non sintetizzate* (*unpooled opinions*): i vettori di probabilità sono in tutto N e ciascuno descrive lo stato di conoscenza di un esperto che fa parte di una squadra di N persone. Più impegnativamente si può sostenere che le probabilità sono *intrinsecamente imprecise*, e dunque sono intervalli e non numeri; è questo il punto di vista che avremo in mente nel seguito.

D'ora in avanti supporremo che $\mathcal{P}$, oltre che compatto, sia anche convesso.

Esercizio. Richiedere che $\mathcal{P}$ sia convesso non è restrittivo. E infatti sia $\mathcal{P}$ chiuso, e sia $\mathcal{I}(\mathcal{P})$ il suo *involucro convesso*: $P, Q \in \mathcal{P} \Rightarrow [P, Q] \subseteq \mathcal{I}(\mathcal{P})$. Tecnicamente l'involucro convesso è il più piccolo soprainsieme convesso di $\mathcal{P}$, ottenuto intersecando *tutti* i soprainsiemi convessi di $\mathcal{P}$; se due "punti" P e Q appartengono all'involucro $\mathcal{I}(\mathcal{P})$, all'involucro appartiene l'intero segmento che li congiunge. Il "funzionale" da rendere massimo e minimo su A, ossia la probabilità , è lineare, per cui se P e Q appartengono a $\mathcal{P}$ e se $R \doteq \lambda P + (1-\lambda)Q$ è un generico punto del segmento $[P, Q]$, $0 \leq \lambda \leq 1$, si ha $R(A) = \lambda P(A) + (1-\lambda)Q(A)$. Dunque la R-probabilità di A è intermedia fra la sua P-probabilità e la Q-probabilità. Ne viene che estremizzare su $\mathcal{P}$ o su $\mathcal{I}(\mathcal{P})$ porta a risultati coincidenti.

Esempio. Tipicamente una probabilità intervallare viene individuata definendo $\mathcal{P}$ mediante vincoli lineari come i seguenti: $p_1 + p_2 \geq \frac{1}{2}$ e $p_2 + p_3 \leq \frac{1}{2}$, $K = 3$. A $\mathcal{P}$ appartengono, ad esempio, $(1, 0, 0)$, $(\frac{1}{2}, 0, \frac{1}{2})$ e $(\frac{1}{2}, \frac{1}{2}, 0)$; grazie alla linearità dei vincoli è facile verificare che $\mathcal{P}$ è convesso. Si potrebbe dimostrare che su universi finiti non è restrittivo limitarsi alla considerazione di famiglie $\mathcal{P}$ definite appunto tramite vincoli lineari. Nel nostro caso si ha ($\square$! aiutati con uno schizzo):
$0 = P_*(\emptyset) = P_*(\{a_2\}) = P_*(\{a_3\}) = P_*(\{a_2, a_3\}) = P^*(\emptyset)$
$\frac{1}{2} = P_*(\{a_1\}) = P_*(\{a_1, a_2\}) = P_*(\{a_1, a_3\}) = P^*(\{a_2\}) = P^*(\{a_3\}) = P^*(\{a_2, a_3\})$
$1 = P_*(\mathcal{X}) = P^*(\{a_1\}) = P^*(\{a_1, a_2\}) = P^*(\{a_1, a_3\}) = P^*(\mathcal{X})$

Si verifica subito ($\square$!) che le probabilità inferiori e le probabilità superiori sono distribuzioni di fiducia; altrettanto semplice è verificare ($\square$!) che le probabilità inferiori sono *superadditive* mentre quelle superiori sono *subadditive*:

$$A \cap B = \emptyset \ \Rightarrow \ P_*(A \cup B) \geq P_*(A) + P_*(B) \ , \ \ P^*(A \cup B) \leq P^*(A) + P^*(B)$$

Inoltre ($\square$!) P_* e P^* sono *duali* l'una dell'altra: $P_*(A) + P^*(\overline{A}) = 1$ (la barra indica la negazione, ossia la complementazione di insieme).

Esercizio. L'implicazione precedente si lascia rafforzare a:

$$A \cap B = \emptyset \ \Rightarrow$$

$$\Rightarrow \ P_*(A) + P_*(B) \leq P_*(A \cup B) \leq P_*(A) + P^*(B) \leq P^*(A \cup B) \leq P^*(A) + P^*(B)$$

Per dimostrare ad esempio la seconda disuguaglianza, prendi $Q = \arg P_*(A)$, ossia $Q(A) = P_*(A)$, $Q \in \mathcal{P}$. Si ha allora $Q(A \cup B) = P_*(A) + Q(B)$ (Q è una normale probabilità bayesiana).

Un *corpus* di attestazioni m (un *body of evidence* o BOE) identifica un insieme $\mathcal{P} = \mathcal{P}(m)$ di vettori di probabilità P *soggiacenti* se si impone che per ogni sottinsieme A dell'universo $\mathcal{X}$ sia:

$$\mathrm{Bel}(A) \leq P(A) \leq \mathrm{Pl}(A)$$

Al BOE vacuo, per fare un esempio, sono soggiacenti tutti i vettori di probabilità. In ogni caso $\mathcal{P}(m)$ è compatto e convesso ($\square$!).

Esercizio. L'insieme dei vettori di probabilità P per cui $P(A) \geq \mathrm{Bel}(A)$ $\forall A$ non è più ampio di $\mathcal{P}(m)$ (usa la dualità). Allo stesso modo: l'insieme dei vettori di probabilità P per cui $P(A) \leq \mathrm{Pl}(A)$ non è più ampio di $\mathcal{P}(m)$.

A probabilità soggiacenti si arriva quando si tenta di approssimare un BOE mediante probabilità bayesiane, come ora proveremo a fare.

Approssimazioni bayesiane dei BOE.
"Parcellizziamo" la massa $m(F)$ dei focali sui singoletti che li costituiscono mediante "parcelle" $m_i(F) \geq 0$: $\sum_{a_i \in F} m_i(F) = m(F)$. Assegniamo ad ogni singoletto la somma delle parcelle che "eredita" dai focali cui appartiene: $P(a_i) \doteq \sum_{F: a_i \in F} m_i(F)$. Per inciso, una "parcellizzazione" uniforme era già stata usata per trovare il massimo della pseudo-entropia $D(m)$. Si ha:
P è una probabilità bayesiana (□ !): diremo che essa *approssima* il BOE.

Beninteso, stante l'arbitrarietà delle parcellizzazioni le approssimazioni bayesiane sono infinite non appena ci siano focali di numerosità almeno 2. Esse sono tutte soggiacenti a m, e dunque contenute in $\mathcal{P}(m)$, come subito si verifica:
Se P approssima m, $\forall A$ è $\mathrm{Bel}(A) \leq P(A) \leq \mathrm{Pl}(A)$
Grazie alla dualità basta dimostrare la prima disuguaglianza; si ha:
$P(A) = \sum_F \sum_{a_i \in A} m_i(F) \geq \sum_{F \subseteq A} m(F)$

Le probabilità bayesiane approssimate sono dunque intermedie fra le credibilità e le plausibilità: in realtà, $\forall A$ esiste un'approssimazione P per cui $P(A) = \mathrm{Bel}(A)$ e un'approssimazione P' per cui $P'(A) = \mathrm{Pl}(A)$. Per dimostrare la prima uguaglianza usa una qualunque parcellizzazione per la quale, se il focale F interseca A, le parcelle interne ad A sono tutte nulle. Per dimostrare la seconda uguaglianza usa una qualunque parcellizzazione per la quale, se il focale F interseca A, le parcelle esterne ad A sono tutte nulle. Dunque, $\forall A$, credibilità e plausibilità si possono ottenere come minimi e massimi su $\mathcal{P}$, anzi sono proprio l'approssimazione bayesiana più piccola e quella più grande:

$$\mathrm{Bel}(A) = \min_{P \in \mathcal{P}(m)} P(A) = P_*(A) \;, \quad \mathrm{Pl}(A) = \max_{P \in \mathcal{P}(m)} P(A) = P^*(A)$$

In questo senso *formale* i BOE sono speciali probabilità intervallari. E' un punto controverso se l'inclusione, oltre che formale, sia anche sostanziale. In ogni caso, l'inclusione è *stretta*, come mostreremo fra poco.

Una curiosa asimmetria. Poniamo per i K singoletti: $\mathrm{bel}_i = \mathrm{Bel}(x_i) \leq \mathrm{pl}_i = \mathrm{Pl}(x_i)$. Ovviamente $\sum_i \mathrm{bel}_i \leq 1$, $\sum_i \mathrm{pl}_i \geq 1$ (□ !). Ebbene, l'eccesso rispetto a 1 di $\sum_i \mathrm{pl}_i$ batte il deficit di $\sum_i \mathrm{bel}_i$:

$$\sum_i \mathrm{bel}_i + \sum_i \mathrm{pl}_i \geq 2$$

E infatti: nella prima sommatoria compaiono solo i focali-singoletti; nella seconda sommatoria ciascun focale compare tante volte quant'è la sua numerosità (quanti sono i singoletti che lo compongono). Dunque nella somma delle due sommatorie ciascun focale compare almeno due volte. Per inciso: si ha uguaglianza se e solo se la numerosità dei focali è al più 2.

L'esempio che segue mostra che le probabilità intervallari non soffrono dell'asimmetria appena segnalata, ed è sufficiente per concludere che la capacità rappresentativa dei BOE è *strettamente inferiore* a quella delle probabilità intervallari.

Esempio. Definiamo $\mathcal{P}$ tramite i K vincoli $1/K \leq p_1 \leq 1/K + \epsilon$, $1/K - \epsilon \leq p_i \leq 1/K$, con $2 \leq i \leq K$, $0 < \epsilon \leq 1/K$. Fra i vettori di $\mathcal{P}$ c'è ad esempio $P = (1/K + \epsilon, 1/K - \epsilon, 1/K, 1/K, \ldots)$. Per le probabilità inferiori e superiori dei singoletti si ha ovviamente $p_{*,1} = 1/K$, $p^*_1 = 1/K + \epsilon$, $p_{*,i} = 1/K - \epsilon$, $p^*_i = 1/K$, $i \geq 2$. Per $K \geq 3$ è allora $\sum_i p_{*,i} + \sum_i p^*_i < 2$, un'asimmetria contraria a quella cui soggiacciono i BOE.

Per concludere, affrontiamo il problema di valutare l'incertezza associata a $\mathcal{P}$. Per una generica probabilità intervallare $\mathcal{P}$, e in particolare per un BOE m, poniamo:

$$H_*(\mathcal{P}) = \min_{P \in \mathcal{P}} H(P) \; , \quad H^*(\mathcal{P}) = \max_{P \in \mathcal{P}} H(P)$$

L'intervallo $[H_*(\mathcal{P}), H^*(\mathcal{P})]$ include le "incertezze alla Shannon" di tutte le probabilità di $\mathcal{P}$, e lo si può assumere come un misura dell'incertezza globale della probabilità intervallare, o, al caso, del BOE. La sua ampiezza dà un indicazione della non-specificità di $\mathcal{P}$ e si riduce a 0 per le probabilità bayesiane. Ricorrere a un intervallo invece che a un singolo scalare ci sembra abbastanza naturale in una teoria di "tipo intervallare", com'è la teoria delle probabilità imprecise o la teoria dell'attestabilità.

Esercizio. $H_*(m) = H^*(m)$ se e solo se $N(m) = 0$, ossia se e solo se l'assegnazione m è bayesiana. Cenno: non esistono segmenti $[P, Q]$ sui quali la funzione entropia, che è *strettamente* concava, sia costante.

Esercizio. Esistono BOE *non vacui* per cui $[H_*(m), H^*(m)] = [0, \log K]$? Cenno: usa un vettore di possibilità (un BOE a focali annidati) con tutte le K componenti $\geq 1/K$.

Complementi sulle probabilità intervallari semplici.

Esercizio. Dimostriamo un'ulteriore disuguaglianze per le credibilità e per le plausibilità dei singoletti. Abbiamo già detto che è $\sum_i \text{bel}_i \leq 1$, $\sum_i \text{pl}_i \geq 1$. Ebbene, queste disuguaglianze si lasciano rafforzare:

$$\sum_{i:i\neq j} \text{bel}_i + \text{pl}_j \leq 1 \quad \text{e}$$
$$\sum_{i:i\neq j} \text{pl}_i + \text{bel}_j \geq 1 \;, \quad 1 \leq j \leq K$$

□ ! Usa la definizione di credibilità e plausibilità: nelle disuguaglianze in alto nessun focale è presente a sinistra più di una volta; in basso ciascun focale è presente almeno una volta; rammenta che il peso totale dei focali è esattamente 1.

Una probabilità intervallare si dice *semplice* se è definita da K vincoli sui singoletti

$$a_i \leq p_i \leq b_i \;, \quad 1 \leq i \leq K$$

con $0 \leq a_i \leq b_i \leq 1$. Beninteso, ipotizziamo sempre che sia $\sum_i a_i \leq 1$, $\sum_i b_i \geq 1$: ciò assicura che l'insieme dei vettori di probabilità coinvolto non è vuoto.

Posto come sopra per le probabilità intervallari dei singoletti $p_{*,i} = P_*(a_i)$, $p_i^* = P^*(a_i)$, si ha ovviamente $a_i \leq p_{*,i} \leq p_i^* \leq b_i$. Purtroppo le due disuguaglianze esterne possono essere strette, come mostra questo banale esempio: $0 \leq p_1 \leq 1$, $p_2 = \frac{1}{2}$, $K = 2$. Se invece $p_{*,i} = a_i$, $p_i^* = b_i$, $1 \leq i \leq K$, l'assegnazione delle probabilità semplici si dice *propria*.

Esercizio. Si ha:

$$p_{*,i} \geq \max[a_i, 1 - \sum_{j\neq i} b_j] \quad , \quad p_i^* \leq \min[b_i, 1 - \sum_{j\neq i} a_j]$$

Per dualità basta dimostrare le affermazioni sulle probabilità inferiori. Si ha $p_{*,i} = 1 - P^*\big(\mathcal{X} - \{x_i\}\big) \geq 1 - \sum_{j\neq i} p_i^* \geq 1 - \sum_{j\neq i} b_j$.

In realtà, come si potrebbe dimostrare, le $2K$ disuguaglianze nell'enunciato dell'esercizio sono addirittura delle uguaglianze, ciò che consente di sostituire una qualunque assegnazione semplice con un'assegnazione equivalente che è anche propria. Basta "aggiornare" i valori che compaiono nei vincoli mediante la regola seguente, dove alla destra delle frecce compaiono i valori "vecchi", a sinistra quelli "nuovi", aggiornati:

$$a_i \to \max[a_i, 1 - \sum_{j\neq i} b_j] \quad , \quad b_i \to \min[b_i, 1 - \sum_{j\neq i} a_j]$$

Un'assegnazione propria è ad esempio quella che è stata usata sopra per dimostrare che la capacità rappresentativa dei BOE è strettamente inferiore a quella delle probabilità intervallari. Lo stesso esempio mostra che ci sono probabilità *semplici* che non possono venir simulate dai BOE. Rinviando alla letteratura specializzata per la dimostrazione (Sgarro, 1998), ci limitiamo a enunciare un criterio che mostra come ci siano numerosi BOE che non possono venir simulati dalle probabilità semplici. Ciò vanifica la speranza che le probabilità semplici costituiscano una sotto-teoria delle probabilità imprecise pratica ma nello stesso tempo abbastanza potente da poter rappresentare tutti i BOE.

Criterio di semplicità: Un BOE può essere rappresentato da probabilità intervallari semplici sse *non* si verifica il seguente fatto: esistono due focali F e G con $F \geq 2$, $G - F \geq 2$.