SOFT COMPUTING
e LOGICA FUZZY

Elaborazione degli stati di conoscenza incompleti: incertezza e vaghezza
Probabilità eterodosse, logiche multivalenti
Calcolo flessibile e aritmetica sfocata

Incomplete information processing: uncertainty and vagueness
Unorthodox probabilities, multivalued logics
Soft computing and fuzzy arithmetics

Andrea Sgarro, Laura Franzoi - DMG, Trieste

Non mi sembra molto logico, obiettò la Volpe

Be', non è poi del tutto illogico, si difese il Gatto

Sulla scala da zero a uno il grado di logicità della vostra affermazione è zero virgola sessantadue centesimi, s'intromise Alice con aria saccente, ma subito si corresse: più o meno

Elucubrazioni prive di ogni scientificità! tagliò corto la Regina di Picche, e con un gesto severo si avvolse nel suo tradizionale mantello bianco e nero, senza la menoma sfumatura di grigio

Alice altrove

Chapter 1

A mo' d'introduzione

1.1 Dilemmi docimologici: un brano di 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 uscì 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[1] 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

[1] In alcune scuole sette e tre quarti: troppo *fuzzy* persini per i *fuzzisti!*

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, diciamo in gradi Celsius, 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?)

Lo strumento tradizionale per rappresentare e gestire le conoscenze incomplete non sono gli insiemi eterodossi, bensì la probablità ortodossa. E' da lì che partiremo, per passar subito alle "probabilità eterodosse".

Chapter 2

Probabilità eterodosse

2.1 Probabilità ortodosse (bayesiane) e distribuzioni di consenso

Nella terminologia probabilistica abituale i K elementi dell'universo di riferimento, o dello *spazio campionario*, o dell'*alfabeto* $\mathcal{X} = \{x_1, x_2, ..., x_K\}$ si chiamano *eventi elementari* (o *simboli, lettere*) 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. Di seguito ci limiteremo al caso finito: nella gestione dell'incertezza, ortodossa o eterodossa che sia, è sul finito che ci sono le idee pregnanti; ad ogni buon conto per le applicazioni ai sistemi esperti di solito basta.
>
> Diremo *evento vuoto* $\emptyset$, *evento pieno* $\mathcal{X}$ ed eviteremo apposta i termini correnti, evento impossibile, evento certo, che creerebbero ambiguità.
>
> Le probabilità ortodosse hanno *almeno* (ma si veda l'Appendice) 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 si usa fare 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:

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

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

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

$\square$: l'assioma $P(\emptyset) = 0$ è sovrabbondante (poni $B = \emptyset$ nell'ultimo assioma).

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

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

(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
$\Phi(A \cap B) \leq \min[\Phi(A), \Phi(B)]$

$\square$: 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; $\overline{\lambda} \doteq 1 - \lambda$, $\lambda \in [0, 1]$. Le probabilità bayesiane sono *autoduali*: $\Phi^* = \Phi$, ed è per questo che di probabilità duale nel calcolo delle probabilità "ortodosso" non si parla mai. Ulteriori considerazioni sulle distribuzioni di fiducia sono posposte alla fine della prossima sezione.

> Il lettore obietterà a ragione che le distribuzioni di fiducia sono un concetto astratto, ma noi le adoperereremo solo come "cornice di riferimento". Nella letteratura sono chiamate, ciò che è meno intuitivo ma più paludato, *capacità di Choquet*, oppure, ed è fuorviante, *misure sfocate*, dove il temine "misura" viene usato nel senso dell'analisi e di sfocato non c'è granché. Per noi la misura di una distribuzione di fiducia è uno scalare, come nel caso della famosa *entropia di Shannon*, di cui al capitolo?? e della sfocatezza parleremo diffusamente più avanti.

2.2 Possibilità e necessità

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 già orientato alla logica, e che segnano, come via via argomenteremo, un punto di sutura fra teorie "logiciste", come quella del prossimo capitolo, e teorie "probabilizzanti" (*probability-like*), se ci si passa il neologismo.

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", suona all'ingrosso: un'evento è possibile sse (sse $\doteq$ se e solo se) 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):

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

Le proprietà e gli esempi che seguono dovrebbero via via convincere il lettore che la scelta di due termini "pesanti" come possibilità e necessità è di fatto adeguata.

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

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

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 [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), $[0, 1]$ (ignoranza totale), $[0, \alpha]$ (contingenza della possibilità), $[\alpha, 1]$ (contingenza della necessità); 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); $\Pi(\overline{A}) = 0 \Rightarrow N(A) = 1$.

Dualmente, poiché $0 = N(\emptyset) = N(A \cap \overline{A}) = \min[N(A), N(\overline{A})]$, almeno uno dei due 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 $\Pi(A) = 1$. QED

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

$\mathcal{X} = \{x_1, x_2, ..., x_K\}$
$\Pi = (\pi_1, \pi_2, ..., \pi_K)$ 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 ($\Box$, usa la dualità); rammentiamo che il co-singoletto $\{\overline{x}\}$ è composto da *tutti* gli elementi dell'universo $\mathcal{X}$ (dello *spazio campionario* $\mathcal{X}$), eccettuato il singoletto x.

Il legame fra la distribuzione e il vettore è così stretto che non crea ambiguità usare per entrambi lo stesso simbolo Π. Sul finito le *probabilità* ortodosse P vengono di solito anch'esse assegnate mediante un vettore di probabilità P con $P(A) = \sum_A p_i$. Formalmente, il calcolo delle possibilità appare dunque come una controparte *maxitiva* del calcolo delle probabilità, che è invece *additivo* (dove, se si passa al continuo, le somme sono ovviamente sostituite da "somme generalizzate", ossia da serie e integrali definiti).

Lemma della persistenza ($\Box$). Se $A \neq \emptyset$ è un evento di possibilità uguale ad α esiste un singoletto di quella possibilità. Dualmente, se $A \subset \mathcal{X}$ è un evento di necessità uguale ad α esiste un singoletto di possibilità $1 - \alpha$.

Lemma del singoletto. Al più un singoletto a_i può avere necessità (strettamente) positiva; $N(a_i) > 0$ implica $\Pi(a_i) \doteq \pi_i = 1 > \pi_j,\ j \neq i$.
Dimostrazione: E difatti $N(a_i) > 0$ è lo stesso di $1 - \Pi\{\overline{a_i}\} > 0$, ossia $\max_{j:\, j \neq i}$ $\pi_j\ < 1$.

Esempio binario. L'unica informazione che ha il soggetto (o l'*agente*) sia: *l'evento F è possibile, il resto no*; $F \subseteq \mathcal{X}$. Come modellare questo stato di conoscenza "povero", senza fingere di sapere ciò che non si sa? Possiamo assegnare la possibilità 1 ai singoletti (eventi elementari) di F, zero altrove. Gli intervalli $[N(E), \Pi(E)]$ sono dei tre tipi $[0, 0]$ se $E \cap F = \emptyset$ (l'evento E è impossibile), $[1, 1]$ se $E \subseteq F$ (E è necessario), $[0, 1]$ altrimenti (di E non possiamo dir nulla, ne siamo totalmente ignoranti). Con sollievo osserviamo che i risultati che abbiamo trovato tornano bene con i due impegnativi termini che avevamo scelto, possibilità e necessità.

Esempio: come "degradare" una probabilità. Dato un vettore di probabilità P sull'universo di riferimanto, binarizziamolo codificando le probabilità elementari positive con 1 e conservando le probabilità elementari 0: il vettore di probabilità P viene sostituito dal vettore di possibiltà Π che è decisamente meno informativo. Il degrado (la perdita d'informazione) ci ha ricondotti al caso precedente.

Esempio psicologico. Passiamo a un esempio non binario, in cui "numerizziamo" quattro stati d'animo: piena fiducia nella possibilità, qualche perplessità, incredulità però magari sì e piena convinzione dell'impossibilità potrebbero venir *numerizzati* mediante le possibilità 1, 2/3, 1/3, 0. Operando su questi quattro numeri con le operazioni consentite non si trova nessun nuovo valore numerico, e dunque nessuna nuova assegnazione numerica che corrisponda a valutazione *linguistica, verbale* diverse dalle quattro precedenti. Stiamo dando i numeri, sia in senso letterale sia in senso traslato? Ne dovremo riparlere nel prossimo capitolo.

Esempio delle uova:
Per chiarire la differenza fra i concetti di probabilità e di possibilità faremo ricorso a un esempio di Zadeh. Si pensi ai sei eventi "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.

$$\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \geq 5 \\ 1 & 1 & .9 & .5 & .1 & 0 \\ .1 & .8 & .1 & 0 & 0 & 0 \end{array}$$

E' del tutto possibile che Pierino mangi due uova, $\text{Poss}\{x = 2\} = 0,9$, 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 $\text{Poss}\{x = 0\} = 1$ mentre $\text{Prob}\{x = 0\} = 0,1$. 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".

> Come giá argomentato in ?? , 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.

$\Box$: un'assegnazione congiunta di probabilità e possibilità non è sempre coerente se è $p_i \leq \pi_i$ soltanto per i K singoletti; si pensi ad esempio a $P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $\Pi = (1, \frac{1}{4}, \frac{1}{4})$. Bastano tuttavia K controlli, come mostra il criterio che segue:

Criterio $\Box$: 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 ...).

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

> 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, hanno insomma la stessa *capacità rappresentativa*. 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". Se x è un singoletto, una distribuzione di fiducia si dice *deterministica su* x se è bayesiana deterministica con $\text{Prob}\{x\} = 1$.

Teorema. Una distribuzione di probabilità P e una distribuzione di possibilità Π coincidono su ogni evento A se e solo se esiste x tale che entrambe siano 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. Se $\forall A, P(A) = \Pi(A)$, si ha $P(\mathcal{X}) = \Pi(\mathcal{X})$, ossia $\sum_i p_i = 1$, e anche $\max_i p_i = 1$, per cui un solo addendo p_i è non nullo, e vale proprio 1.
La seconda affermazione scende dalla prima per dualità, poiché le probabilità bayesiane sono *autoduali*: $P^*(a) \doteq 1 - P(\overline{A}) = P(A)$.
Se $\forall A, N(A) = \Pi(A)$ si ha in particolare $\forall x N(x) = \Pi(x)$. Basta usare il lemma del singoletto.

Un addendum stellato sulle distribuzioni di fiducia.

Se $\Phi \doteq \lambda\Pi + \overline{\lambda}N$, $\lambda \in [0,1]$, $\overline{\lambda} \doteq 1-\lambda$, $N = \Pi^*$, si dimostri $\square$ che Φ domina la sua duale Φ^* se $\lambda > 1/2$, ne è dominata se $\lambda < 1/2$, mentre se $\lambda = 1/2$ la Φ è autoduale. Ciò ci dà subito il destro di dare l'esempio di una distribuzione di fiducia autoduale che *non* sia una probabilità bayesiana. Vedremo fra poco che su universi *binari* qualunque disribuzione di fiducia è simulabile nell'ambito dell'*evidence theory*, per cui nel seguito dell'addendo supporremo senza reale restrizione che l'universo sia almeno ternario.

Prendiamo due eventi (due sottinsiemi) A e B disgiunti, non vuoti e non esaurienti. Senza restrizione effettiva, supponiamo che tutti i singoletti (e dunque ogni evento non vuoto) abbiano possibilità strettamente positive e che sia $\Pi(B) \leq \Pi(A)$; ne viene $\Pi(\overline{B}) = 1$, $\Pi(A \cup B) = \Pi(A)$. Se la Φ fosse una probabilità additivita, $\Phi(A \cup B) = \Phi(A) + \Phi(B)$, ricordando la sua definizione in temini di Π, dovrebbe essere: $\Pi(\overline{A}) = \Pi(B) + \Pi(\overline{A \cup B})$. Ora, $\overline{A}$ è unione (disgiunta) di B e di $\overline{A \cup B}$, e dunque la sua possibilità è uguale ad almeno una delle due possibilità che figurano al secondo membro della presunta uguaglianza, possibilità che tuttavia sono entrambe strettamente positive.

Visti gli esempi finora dati e quelli che daremo più avanti, il lettore potrebbe essere indotto a pensare che una distribuzione di fiducia sia vincolata a dominare o a venir dominata dalla sua duale, e che dunque sia atta a costruire una teoria di tipo intervallare. Purtroppo non è così, come ora mostriamo, il che sottolinea l'eccessiva genericità di questo strumento. Un lemmino di premessa per tre distribuzioni di fiducia del tutto arbitrarie: se Φ è ottenuta come combinazione convessa di Φ_1 e Φ_2, $\Phi = \lambda\Phi_1 + \overline{\lambda}\Phi_2$, la duale di Φ è combinazione convessa delle duali di Φ_1 e di Φ_2 con gli stessi coefficienti, $\Phi^* = \lambda\Phi_1^* + \overline{\lambda}\Phi_2^*$. Sull'universo ternario $\{a,b,c\}$ consideriamo ora i due vettori di possibilità $\Pi_1 = (1,1,0)$ e $\Pi_2 = (0,1,1)$, e le corrispondenti distribuzioni di possibilità e di necessità Π_1, Π_2, $N_1 = \Pi_1^*$ e $N_2 = \Pi_2^*$. Per $\lambda \in]0,1[$ sia Φ combinazione convessa di Π_1 e N_2, e dunque sia Φ^* combinazione convessa di N_1 e Π_2. Si ha: $\Phi(a) = \lambda > \Phi^*(a) = 0$, ma, dualmente e simmetricamente, $\Phi^*(c) = \overline{\lambda} > \Phi(c) = 0$.

2.3 Teoria dell'attestabilità (*evidence theory*)

Sul piano epistemico (in greco *episteme* = conoscenza), la probabilità dell'evento A o anche $\{X \in A\}$, $A \subseteq \mathcal{X}$, viene interpretata come il *grado di fiducia* che l'agente umano ha nella realizzazione di A (X è una *variabile* o un *attributo aleatorio*: si veda l'appendice).

> 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*; sottolineiamo che due focali possono intersecarsi. 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*):

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

Non è m, ma sono questi due numeri, casomai, che fanno le veci della probabilità di A, Prob(A): in effetti, alcuni vedono l'intervallo [Bel(A),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 $\mathcal{X}$ (né vuoto né pieno) 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. Infine, un BOE si chiama *non dogmatico* quando $m(\mathcal{X}) \neq 0$.

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, "passando al quoziente") 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 $\Box$ 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 bacchettoni - 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 contraddittorie: la qualità dello smalto lo faceva propendere per il sì, 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, da cui la prima scende per dualità.
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 *insiemi iniziali* del tipo $F_i \doteq \{x_1, x_2, ..., x_i\}$ (ossia "teste"), ponendo $m(F_i) \doteq m\{x_1, x_2, ..., x_i\} = \pi_i - \pi_{i+1}$, $(\pi_{K+1} \doteq 0)$; m individua effettivamente un BOE perché $\sum_F m(F) = \pi_1 = 1$. Attenzione: non tutti i K insiemi iniziali sono necessariamente anche insiemi focali, per qualche i potrebbe ben essere $m(F_i) = m(\{x_1, x_2, ..., x_i\}) = 0$. Sia A un generico evento e sia x_j il suo primo elemento, cioè quello di indice più basso; si ha:
$\mathrm{Pl}(A) = \sum_{i:\, i \geq j} m(F_i) = \sum_{i:\, i \geq j} (\pi_i - \pi_{i+1}) = \pi_j = \max_{x_i \in A} \pi_i = \Pi(A)$
Viceversa, dato un BOE consonante, possiamo sempre supporre, riordinando al caso l'insieme universo $\mathcal{X}$, che i suoi focali siano insiemi iniziali (attenzione di nuovo: non tutti i K insiemi iniziali sono necessariamente anche insiemi focali). Definiamo un vettore di possibilità $\Pi = (\pi_1 = 1 \geq \pi_2 \geq \ldots \geq \pi_K)$ ponendo $\pi_j \doteq \sum_{i:\, i \geq j} m(F_i)$, $1 \leq j \leq K$. Sia A un evento e sia di nuovo x_j il suo primo elemento. Si ha $\square$ $\mathrm{Pl}(A) = \pi_j$. I focali F_i sono insiemi iniziali, e dunque $x_j \in F_i$ sse $i \geq j$, per cui
$\mathrm{Pl}(A) \doteq \sum_{i: A \cap F_i \neq \emptyset} m(F_i) = \sum_{i: i \geq j} m(F_i) \doteq \pi_j = \Pi(A)$

Esercizio. I BOE, a differenza delle distribuzioni di possibilità, godono della proprietà della convessità ($\square$: pensa ai focali e all'unione delle due famiglie dei focali; se un focale è a comune combinane i pesi). Quand'è che una combinazione convessa di due possibilità è ancora una possibilità?

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 $A \neq \emptyset$ (!).

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 $\mathrm{Pl}(A)$.

Una digressione stellata.
Su un universo binario, diciamo $\{a, b\}$, qualunque distribuzione di fiducia Φ è simulabile nell'ambito dell'*evidence theory.* Se $\Phi(a)+\Phi(b) \leq 1$, basterà ricorrere alla corrsipondente probabilità (eventualmente) incompleta; in tal caso la fiducia è una credibilità (una probabilità puntuale se $\Phi(a)+\Phi(b) = 1$). Se $\Phi(a)+\Phi(b) > 1$, basterà passare alla duale, per la quale è $\Phi^*(a) + \Phi^*(b) < 1$; in tal caso la fiducia Φ è invece una plausibilità.

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

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

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

Formula insiemistica di Möbius: Sia $g : 2^{\mathcal{X}} \to \Re$, e sia $f(A) \doteq \sum_{B \subseteq A} g(B)$. 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:
$\sum_{B:B \subseteq A}(-1)^{|A-B|}f(B) =$
$= (-1)^{|A|} \sum_{B:B \subseteq A}(-1)^{|B|}f(B)$
$= (-1)^{|A|} \sum_{B:B \subseteq A}(-1)^{|B|} \sum_{C:C \subseteq B} g(C)$
$= (-1)^{|A|} \sum_{C:C \subseteq A} g(C) \sum_{B:C \subseteq B \subseteq A}(-1)^{|B|} = (-1)^{|A|}g(A)(-1)^{|A|} = g(A)$.

Nella formula "filologica" il reticolo è quello degli interi positivi n parzialmente ordinati per *divisibilità*, dove $3 \prec 6$, $3 \preceq 3$ ma $3 \not\preceq 5$. Le dimostrazioni $\Box$ ricalcano le precedenti. La funzione di Möbius è definita da $\mu(n) = (-1)^r$, dove $n \geq 2$ é uguale al prodotto di r numeri primi *distinti* (n è esente da quadrati, o *square-free*); $\mu(1) \doteq 1$, altrimenti $\mu(n) \doteq 0$. Le sommatorie che seguono vanno estese a tutti i divisori d di n, n incluso; $n \geq 1$:

Lemma. Per $n \geq 2$ la somma $\sum_{d;\, d \preceq n} \mu(d)$ è nulla; vale 1 per $n = 1$.

Formula di Möbius: Sia $g : \mathbf{N} \to \Re$, e sia $f(n) \doteq \sum_{d;\, d \preceq n} g(d)$.
Allora: $g(n) = \sum_{d;\, d \preceq n} \mu(\frac{n}{d}) f(d) = \sum_{d;\, d \preceq n} \mu(d) f(\frac{n}{d})$.

Nella letteratura esistono formule più generali su strutture d'ordine "astratte" di cui le due formule di Möbius, quella filologica e quella insiemistica, sono casi particolari.

2.4 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: $m_2\{\mathrm{Ci}, \mathrm{Cl}, \mathrm{Pa}\} = .9$, $m_2\{\mathcal{X}\} = .\infty$. 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: $\mathrm{Bel}_2(\text{cirrosi}) = 0$, $\mathrm{Pl}_2(\text{cirrosi}) = 1$, $\mathrm{Bel}_2(\text{colestasi intraepatica}) = 0$, $\mathrm{Pl}_2(\text{colestasi intraepatica}) = 1$. 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: $m\{\mathrm{Ci}\} = .72$, $m\{\mathrm{Ep}, \mathrm{Ci}\} = .08$, $m\{\mathrm{Ca}, \mathrm{Pa}\} = .18 + .02 = .2$. 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 prima via d'uscita è la regola non-dogmatica che "regala" la massa avanzata $m(\emptyset)$ alla cornice $\mathcal{X}$ in maniera del tutto non impegnativa (*non-committal*), portando a un BOE *non-dogmatico* che ha la cornice di riferimento $\mathcal{X}$ fra i suoi focali. 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 ed esclusivamente universi chiusi: $m(\emptyset) = 0$. Facciamo notare che ammettere $m(\emptyset) \neq 0$ fa uscire dalla teoria della misura come la conosciamo e introduce problemi tecnici (matematici) da far tremare i polsi.

Un BOE consonante e oggettivo.
Come hanno già mostrato le probabilità incomplete, l'*osservabilità incompleta (parziale)* degli esperimenti consente di introdurre BOE oggettivi, empirici, statistici. Segue un esempio di tipo possibilistico, ossia con focali a catena. Abbiamo due dadi equi diciamo a quattro facce marcate 1,2,3,4 (la loro equità può venire statisticamente convalidata), uno bianco con i punti in nero, uno nero con i punti in bianco: è *solo* il dado bianco che ci interessa. Il dado bianco determina a che livello, 1, 2, 3 o 4, viene riempito di liquido rosso un contenitore trasparente. Prima di farlo vedere all'osservatore viene lanciato il dado nero e il contenitore viene rabboccato, se serve, fino al livello indicato dal dado nero (che intesta le righe):

	1	2	3	4
1	1	≤ 2	≤ 3	≤ 4
2	≤ 2	≤ 2	≤ 3	≤ 4
3	≤ 3	≤ 3	≤ 3	≤ 4
4	≤ 4	≤ 4	≤ 4	≤ 4

I focali sono $\{1\}, \{1,2\}, \{1,2,3\}, \{1,2,3,4\} = \mathcal{X}$ di peso $m = \frac{1}{16}, \frac{3}{16}, \frac{5}{16}, \frac{7}{16}$ rispettivamente ($\square$!). Si noti che l'osservatore, dopo una lunga sperimentazione, potrebbe dare una stima statistica di questi pesi (di queste probabilità su $2^{\mathcal{X}}$). Il vettore di possibilità corrispondente è $\Pi = (1, \frac{15}{16}, \frac{12}{16}, \frac{7}{16})$ (!). Per $A = \{2,4\}$ si ha $\text{Bel}(A) = 0$, $\text{Pl}(A) = \frac{15}{16}$ (i focali che contribuiscono alla plausibilità sono quelli in cui si osserva il liquido a livello 2, 3 o 4).

I probabilisti ortodossi, puntuali e non intervallari, modellerebbero la situazione tramite una coppia aleatoria di variabili dipendenti: il punteggio X che esce sul dado bianco e il livello Y che può venir osservato dopo il rabbocco. Nel linguaggio dei teorici dell'informazione: l'input X viene osservato al di là di un canale *disturbato* da "rumore" stocastico (*noisy channel*) il cui output è Y. Invece di ricorrere a credibilità e plausibilità si direbbe che è 0 la probabilità di un osservazione che *implichi* A e che è $15/16$ la probabilità di un'osservazione che sia compatibile con A, che non implichi la sua negazione. Nell'ottica dei BOE la variabile X non è osservabile sia pure tramite una sua versione disturbata, l'"atomo" dell'osservazione è un focale.

2.5 Probabilità imprecise (intervallari)

Dal punto di vista tecnico, le *probabilità intervallari* fanno anch'esse parte della *teoria delle capacità* proposta da Gustave Choquet negli anni 50 del secolo scorso 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 perfino 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. Imporre la chiusura di $\mathcal{P}$ serve essenzialmente a evitare il fastidio di scrivere *inf* e *sup* (estremo inferiore e superiore) invece di *min* e *max*; oltre che chiuso $\mathcal{P}$ è anche *compatto*, visto che *qualunque* insieme di vettori di probabilità sull'universo $\mathcal{X}$ è limitato in $\Re^{|\mathcal{X}|}$. Se $\mathcal{P}$ non fosse chiuso, lo sostituiremo senza pregiudizio con la sua *chiusura*, vale a dire con il piu piccolo chiuso che lo contiene (con l'intersezione di tutti i chiusi che lo contengono: si può agevolmente dimostrare che quest'intersezione è essa stessa chiusa). 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. □

> *Divagazione filosofica*: La maggior precisione è comunque più informativa, più auspicabile, più sensata? Oppure esistono situazioni *intrinsecamente* imprecise? Quando è *esattamente* scoppiata la prima guerra mondiale? D'accordo, nel 1914, ma vediamo di essere più precisi. In quale mese? Anzi, in quale giorno e a quale ora? E sapere a quale minuto di quell'ora, o magari a quale secondo di quel minuto davvero vi interessa?

D'ora in avanti supporremo che $\mathcal{P}$, oltre che compatto, sia anche *convesso*. Richiedere che $\mathcal{P}$ sia convesso *non* è restrittivo. E infatti sia $\mathcal{P}$ chiuso, e sia $\bar{\mathcal{P}}$ il suo *involucro convesso*: $P, Q \in \mathcal{P} \Rightarrow [P, Q] \subseteq \bar{\mathcal{P}}$, dove $[P, Q]$ è il segmento $|\mathcal{X}|$-dimensionale di estremi P e Q. 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 $\bar{\mathcal{P}}$, all'involucro appartiene l'intero segmento che li congiunge. (Si può dimostrare $\square$ che l'intersezione di un qualsivoglia numero, finito o infinito, di insiemi convessi lo è ancora.) Se P e Q appartengono a $\mathcal{P}$ e se $R \doteq \lambda P + (1-\lambda)Q$ è un generico punto del segmento $[P, Q]$, $0 \leq \lambda \leq 1$, il "funzionale" da rendere massimo e minimo su A, ossia la probabilità $R(A) = \lambda P(A) + (1-\lambda)Q(A)$, è lineare in λ. Dunque la R-probabilità di A è intermedia fra la sua P-probabilità e la Q-probabilità: ne viene che estremizzare (massimizzare o minimizzare) su $\mathcal{P}$ o su $\bar{\mathcal{P}}$ 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. In effetti basta cercare massimi e minimi sulla frontiera lineare, anzi basta cercarli sulle *cuspidi*, ossia sui punti d'angolo della frontiera, che sono in numero finito ($\square$! aiutati con uno schizzo su $\Re^2$, eliminando $p_3 = 1 - p_1 - p_2$; usa argomentazioni simili a quella che ci ha appena permesso di dichiarare non restrittiva la convessità). Si ha:

$0 = P_*(\emptyset) = P_*(\{a_2\}) = P_*(\{a_3\}) = P_*(\{a_2, a_3\}) = P^*(\emptyset)$
$\frac{1}{2} = P_*(\{a_1\}) = P_*(\{a_1, a_2\}) = P_*(\{a_1, a_3\}) = P^*(\{a_2\}) = P^*(\{a_3\}) = P^*(\{a_2, a_3\})$
$1 = P_*(\mathcal{X}) = P^*(\{a_1\}) = P^*(\{a_1, a_2\}) = P^*(\{a_1, a_3\}) = P^*(\mathcal{X})$

Argomentazione euristica: Si può dimostrare che su universi finiti non è restrittivo limitarsi alla considerazione di famiglie $\mathcal{P}$ definite tramite vincoli lineari, come succede nell'esempio, un'affermazione che ci limiteremo ad *argomentare* euristicamente, senza dimostrarla formalmente. Si pensi a $\mathcal{P}$ in $\Re^2$ o $\Re^3$, per poterlo "vedere". Già sappiamo che i massimi e i minimi si trovano sulla frontiera. Potrebbero essere tantissimi, perchè il numero dei sottoinsiemi (degli eventi) sull'universo $\mathcal{X}$ è esponenziale nella numerosità (o *cardinalità*) $|\mathcal{X}|$, ma sono comunque in numero *finito*. Si pensi ai punti corrispondenti sulla frontiera di $\mathcal{P}$, che sono dunque in numero finito. A $\mathcal{P}$ sostituiremo il suo sottinsieme "lineare" (poliedrico, un *politopo*, per usare un termine tecnico della programmazione lineare) ottenuto a partire dalle sue cuspidi, cioè appunto dai punti di massimo e minimo sulla frontiera di $\mathcal{P}$. Estremizzare sul politopo dà gli stessi risultati che estremizzare su $\mathcal{P}$: quanto argomentato implica che la ricerca dei massimi e dei minimi si può limitare a un controllo finito delle cuspidi del politopo.

Si verifica subito □ che le probabilità inferiori e le probabilità superiori sono distribuzioni di fiducia; altrettanto semplice è verificare □ 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 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 $Q(A \cup B) = P_*(A) + Q(B)$ $\big(Q$ è una normale probabilità bayesiana$\big)$.

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 (□ !).

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: $P(a_i) \doteq \sum_{F: a_i \in F} m_i(F)$.

P è una probabilità bayesiana $\square$: 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 $\sum_i \mathrm{bel}_i \leq 1$, $\sum_i \mathrm{pl}_i \geq 1$ ($\square$!). 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.

Complementi stellati sulle probabilità semplici
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:

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

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

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:

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

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, 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 (Sgarro, 1998). Un BOE può essere rappresentato da probabilità intervallari semplici sse *non* si verifica il seguente fatto: esistono due focali F e G con $F \geq 2$, $G - F \geq 2$.

Chapter 3

Fuzziness = sfocatezza

3.1 L'algebra degli insiemi sfocati

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 $f_A(x_i)$.

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:

$$
\begin{aligned}
&\text{commutatività e associatività} : \; A \cup B = B \cup A\,, \; A \cup (B \cup C) = (A \cup B) \cup C \\
&\text{idempotenza e involuzione} : \; A \cup A = A\,, \; A = \overline{\overline{A}}
\end{aligned}
$$

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 né il principio di non contraddizione né 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, o, al caso, l'estremo superiore).

C'è una stretta parentela fra la teoria della possibilità, vista come forma di logica multivalente (*multi-valued logic*), e la teoria 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 tre 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.
Il valore logico della proposizione {*L'oggetto* x *goda della proprietà* A} *è* μ.

Va da sé che gli insiemi *fuzzy* possono venir definiti anche sulla retta reale $\mathbb{R}$ mediante una funzione $\mu(x) : \mathbb{R} \to [0,1]$. Ad esempio, come vedremo più avanti parlando di aritmetica sfocata, un numero come "circa 3" potrebbe venir modellato (codificato) adeguatamente mediante una funzione lineare a due tratti che sale da 0 a 1 sull'intervallo $[2,3]$ e scende da 1 a 0 sull'intervallo $[3,4]$ (*numero fuzzy triangolare*); un'intervallo come "di sicuro fra 3 e 4, al più fra 2 e 5 potrebbe venir adeguatamente codificato mediante una funzione lineare a tre tratti che sale da 0 a 1 sull'intervallo $[2,3]$ scende da 1 a 0 sull'intervallo $[4,5]$ e rimane costante uguale a 1 sull'intervallo $[3,4]$ (*intervallo fuzzy trapezoidale*).

3.2 Prodotti cartesiani e non-interattività

Un sottinsieme *fuzzy* bidimensionale $A \subseteq \mathcal{U}^2$ o più specificamente $A \subseteq \mathbb{R}^2$ è definito dalla sua funzione caratteristica: $\mu_A : \mathcal{U}^2 \to [0,1]$ ossia $\mu_A : \mathbb{R}^2 \to [0,1]$. Vediamo ora di definire in maniera sensata le due *proiezioni* di A, A_X e A_Y, sull'asse cartesiano orizzontale e su quello verticale rispettivamente, nel caso reale su $\mathbb{R}_X$ e su $\mathbb{R}_Y$. In una situazione nitida un numero reale a appartiene allla proiezione A_X quando *almeno* una coppia (a,y) appartiene ad A, dove l'ascissa è proprio a mentre l'ordinata y è arbitraria: all'operazione logica di *disgiunzione* (*oppure* un primo y *oppure* un secondo y *oppure* indifferentemente qualunque altro $y \in \mathcal{U}$ o al caso $y \in \mathbb{R}$) corrisponde l'operazione aritmetica di *massimo* fra i valori logici, dove sul continuo, quando sia coinvolto $\mathbb{R}$, il massimo va rimpiazzato da un estremo superiore *sup* per mere ragioni tecniche. Definiamo allora le due proiezioni sfocate A_X e A_Y ponendo:

$$\mu_{A_X}(x) \doteq \sup_{y \in \mathbb{R}} \mu_A(x,y) \quad , \quad \mu_{A_Y}(y) \doteq \sup_{x \in \mathbb{R}} \mu_A(x,y)$$

dove al caso $\mathbb{R}$ va sostituito da un generico $\mathcal{U}$ e il *sup* è anche un *max*. L'estensione a insiemi n-dimensionali con $n \geq 3$ è immediata.

Passiamo a definire in maniera sensata il *prodotto cartesiano* $A \times B$ (insieme delle coppi) di due insiemi sfocati A e B. 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* $y \in B$ 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:

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

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 volentieri perchè aiuta la leggibilità delle formule. L'estensione al prodotto di più di due insiemi è immediata. Nell'aritmetica *fuzzy*, come si vedra, i prodotti cartesiani hanno un ruolo fondamentale.

Se dell'insieme bidimensionale A conosciamo solo la proiezione orizzontale A_X e la proiezione verticale A_Y *non* riusciamo □ a ricostruire A in maniera univoca. Una possibile ricostruzione □ potrebbe essere quella di "incollare" le due proiezioni con un minimo: $\mu_{A}(x, y) \doteq \mu_{A_Y}(x) \wedge \mu_{A_Y}(y)$. Se la ricostruzione è corretta diciamo che A è *non-interattivo*: la scelta di un termine così impegnativo, che ricorda (non a caso) l'*indipendenza* del calcolo delle probabilità, si chiarirà parlando dell'impostazione possibilistica dell'aritmetica sfocata, cfr. oltre la sezione??.

3.3 Il principio di estensione e le quantità sfocate

Introdurremo ora 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 $\mathcal{X}$ finito o meno 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:

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

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 $\big[f(A)\big](y) = 0$, com'è naturale.

Veniamo alle quantità *fuzzy*: anticipare a questo punto argomenti di aritmetica *fuzzy* ci consente una brillante applicazione del principio di estensione. L'ambiente di riferimento non è più finito, ma del resto molte delle nozioni delle sezioni precedenti sono generalizzabili senza difficoltà, a partire da quella di (sott)insieme *fuzzy*. 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 appartenenza 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 $a \leq r = s \leq b$ si ritrovano i numeri sfocati triangolari.

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 biunivocitị 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 $y \geq 0$, $B(y) = \max\big[A(-\sqrt{y}), A(+\sqrt{y})\big]$; $\square$. 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 $(-1, 0, +1)$.

3.4 *Relazioni sfocate

Nella matematica *nitida*, quella tradizionale, una *relazione* sull'insieme "universo" $\mathcal{X}$, che per il momento supponiamo finito (e non vuoto), di fatto si "identifica" con un sottinsieme $\mathcal{R}$ del quadrato cartesiano $\mathcal{X}^2$, ossia dell'insieme delle coppie (x, y), $x, y \in \mathcal{X}$. Se la relazione $\mathcal{R}$ coinvolge *due* insiemi nitidi $\mathcal{X}$ e $\mathcal{Y}$, sempre finiti, $\mathcal{R}$ si identifica con un sottinsieme del prodotto cartesiano $\mathcal{X} \times \mathcal{Y}$, ossia dell'insieme delle coppie (x, y), $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Esattamente allo stesso modo si definisce una *relazione sfocata* (*fuzzy relation*, *relation floue*), salvo che il sottinsieme delle coppie $\mathcal{R}$ può stavolta essere *sfocato* (*fuzzy*, *flou*). Concisamente, scriveremo $\mathcal{R}(x, y)$ per indicare sia il grado di appartenenza di (x, y) al sottimsieme bidimensionale $\mathcal{R}$, sia il valore logico della proposizione "va ga" $\{(x, y) \in \mathcal{R}\}$; $\mathcal{R}(x, y) \in [0, 1]$.

Esempio: sia l'universo $\mathcal{X}$ costituito da un certo numero di poeti, e la relazione $\mathcal{R}$ sia: x e y sono contemporanei. Potremmo decidere di porre $\mathcal{R}(x, y) = 1$ se x e y sono nati nello stesso giorno dello stesso mese dello stesso anno, $\mathcal{R}(x, y) = 0.95$

se il giorno è un altro ma il mese e l'anno tornano, $\mathcal{R}(x, y) = 0.9$ se torna solo l'anno, altrimenti $\mathcal{R}(x, y) = 0.5$ se la differenza fra i due anni di nascita è comunque ≤ 30, e se no $\mathcal{R}(x, y) = 0$.

La relazione precedente è simmetrica e riflessiva: in effetti alcune nozioni nitide si generalizzano subito alle relazioni sfocate (nelle definizioni seguenti è omesso il $\forall x, y, z$ e si suppone $\mathcal{X} = \mathcal{Y}$):

$\mathcal{R}$ è *riflessiva* se e solo se $\mathcal{R}(x, x) = 1$

$\mathcal{R}$ è *simmetrica* se e solo se $\mathcal{R}(x, y) = \mathcal{R}(y, x)$

$\mathcal{R}$ è *antisimmetrica* se e solo se $\mathcal{R}(x, y) \neq 0 \ \wedge \ \mathcal{R}(y, x) \neq 0$ implica $x = y$

Equivalentemente l'antisimmetria richiede che sia $\Box$:
$x \neq y \ \wedge \ \mathcal{R}(x, y) \neq 0$ implica $\mathcal{R}(y, x) = 0$.

Più problematica è la *transitività*, che ci costringe a definire preliminarmente la *composizione* $\mathcal{R} \circ \mathcal{T}$ di due relazioni $\mathcal{R}$ e $\mathcal{T}$ applicate nell'ordine, prima $\mathcal{R}$ fra $\mathcal{X}$ e $\mathcal{Z}$ e poi $\mathcal{T}$ fra $\mathcal{Z}$ e $\mathcal{Y}$. Posto $\mathcal{R} \circ \mathcal{T} \doteq \mathcal{K} \subseteq \mathcal{X} \times \mathcal{Z}$, qual è il valore logico "naturale" della proposizione $P_{x,y} \doteq \{(x, y) \in \mathcal{K}\}$, ossia quale dev'essere il grado di appartenenza di (x, y) a $\mathcal{K}$? Ebbene (x, y) appartiene a $\mathcal{K}$ *nella misura in cui* esiste uno z tale che (x, z) appartenga a $\mathcal{R}$ e (z, y) appartenga a $\mathcal{T}$. Tale misura è il grado di verità della proposizione *esiste uno z tale che sia $(x, z) \in \mathcal{R}$ ed anche $(z, y) \in \mathcal{T}$*. Poichè il valore di verità dell'appartenza di uno z prefissato a un tale insieme (il suo grado di appartenenza all'insieme) è $\mathcal{R}(x, z) \wedge \mathcal{T}(z, y)$, il valore di verità della proposizione $P_{x,y}$ è:

$$\mathcal{K}(x, y) \doteq \big[\mathcal{R} \circ \mathcal{T}\big](x, y) \ = \ \vee_{z \in \mathcal{X}} \big[\mathcal{R}(x, z) \wedge \mathcal{T}(z, y)\big]$$

La disgiunzione logica (il massimo, o al caso l'estremo superiore) è applicata a *tutti* gli elementi dell'universo nitido $\mathcal{X}$. Siamo ora in grado di generalizzare la nozione di transitività, se $\mathcal{X} = \mathcal{Y} = \mathcal{Z}$, $\mathcal{R} = \mathcal{T}$.

$\mathcal{R}$ è *transitiva* se e solo se $\forall x, y$ si ha $\mathcal{R}(x, y) \geq \mathcal{R}^2(x, y) \doteq \big[\mathcal{R} \circ \mathcal{R}\big](x, y)$

Si richiede insomma che x sia in relazione con y almeno quanto lo è se si è vincolati a"passare" da x a y tramite un qualunque z dell'universo $\mathcal{X}$. In termini insiemistici $\Box$ la transitività equivale a $\mathcal{R}^2 \subseteq \mathcal{R}$. Le *relazioni d'ordine* sfocate saranno, come nel caso nitido, quelle che sono al contempo riflessive, antisimmetriche e transitive. La relazione dell'esempio poetico *non* è transitiva $\Box$.

3.5 Norme e conorme triangolari

Puntiamo per un po' a una maggior astrazione, come piace ai matematici. Nella letteratura la nostra proposta *standard* di congiunzione, disgiunzione e negazione logiche non è l'unica ad aver corso, anche se, nel caso della negazione, ci sono pochi rivali alla $1-x$, dove $x \in [0,1]$ è un grado di appartenenza oppure un valore logico di verità. A quali proprietà "ineludibili" e dunque "minime" dovrebbero soddisfare congiunzioni, disgiunzioni e negazioni astratte? Cominciamo con l'introdurre gli assiomi delle congiunzioni astratte (*e anche*, AND), che si chiamano *norme triangolari* o T-*norme*. Di seguito $\top$ (e più avanti $\perp$) sono operazioni numeriche ad argomenti x e y; si ha $x,\ y, x\top y \in [0,1]$.

i) $x\top y = y\top x =$ commutatività
ii) $(x\top y)\top z = x\top(y\top z)$ associatività
iii) $y \le u \ \Rightarrow\ x\top y \le x\top u =$ monotonia
iv) $x\top 1 = x$, l'1 è elemento neutro
v) $x\top 0 = 0$, lo 0 è nullifico

L'ultimo assioma in realtà è sovrabbondante, essendo implicato da iii) e iv) ($\square$). Ovviamente, stante la commutatività, gli assiomi da iii) a v) si possono anche scrivere "rovesciati"; non ci sono altri nullifici né altri elementi neutri ($\square$). Passiamo alla disgiunzione astratta (*o anche*, *oppure indifferentemente*, OR), che si chiama *conorma triangolare* o T-*conorma*; in simboli $x\perp y$. Gli assiomi sono analoghi, anzi i primi tre sono identici:

i) $x\perp y = y\perp x =$ commutatività
ii) $(x\perp y)\perp z = x\perp(y\perp z)$ associatività
iii) $y \le u \ \Rightarrow\ x\perp y \le x\top u =$ monotonia
iv) $x\perp 0 = x$, lo 0 è elemento neutro
v) $x\perp 1 = 1$, l'1 è nullifico

Valgono osservazioni analoghe a quelle sulle T-norme, in particolare l'ultimo assioma è sovrabbondante. Seguono gli assiomi della *negazione astratta*, in simboli $\neg x$ o $\bar{x}$:

i) $\neg 0 = 1,\ \neg 1 = 0$
i) $\neg\big(\neg x\big) = x$ involutività
iii) $x \le y \ \Rightarrow\ \neg x \ge \neg y$ antimonotonia

In realtà, stante l'involutività, la seconda parte dell'assioma 1 è sovrabbondante data la prima, o, ciò che è lo stesso, la prima parte è sovrabbondante data la seconda. Data una negazione astratta, T-norme e T-conorme si rivelano anch'esse concetti *duali*, basta imporre (assiomatizzare) la validità delle regole di De Morgan. E infatti ($\square$):

Data una T-norma $\top$ e una negazione astratta $\neg$, $x\perp y \doteq \neg\big(\neg x\top\neg y\big)$ è una T-conorma, *duale* della T-norma $\top$.

Data una T-conorma $\perp$ e una negazione astratta $\neg$, $x\top y \doteq \neg\big(\neg x \perp \neg y\big)$ è una T-norma, *duale* della T-conorma $\top$.

La dualità è involutiva ($\square$), in pratica si assegnano la norma e la negazione e implicitamente si assume che la conorma sia quella duale, o viceversa si assegnano la conorma e la negazione. Inutile dirlo, la scelta più comune è quella standard che abbiamo già fatto, con minimi, massimi e complementi aritmetici a 1. Ma ci sono altre opzioni, per esempio (la negazione d'ora in avanti è sempre quella standard, $\bar{x} = 1 - x$, che garantisce la dualità $\square$):

$\top$	$\perp$	
$x \wedge y$	$x \vee y$	standard
xy	$x + y - xy$	probabilistica
$\big[x + y - 1\big] \vee 0$	$\big[x + y\big] \wedge 1$	di Lukasiewicz

Le norme standard godono di due proprietà estremali, perchè si ha: $x\top y \leq x \wedge y$ e $x \perp y \geq x \vee y$ per qualunque norma $\top$ per qualunque conorma $\perp$; in questo senso il minimo è la norma più grande, mentre il massimo è la conorma più piccola ($\square$, usa la monotonia e l'esistenza dell'elemento neutro per dimostrare $x\top y \leq x$ e $x\top y \leq y$; per dualità basta dimostrare l'affermazione sulla norma). E agli altri estremi? La risposta è data da una coppia duale artificiosa $\square$: la T-norma *drastica* $\top_D$ vale sempre 0 tranne quando gli assiomi la obbligano a valere $x > 0$ perchè l'altro argomento è l'elemento neutro1, mentre la T-conorma *drastica* $\perp_D$ vale sempre 1 tranne quando gli assiomi la obbligano a valere $x < 1$ perchè l'altro argomento è lo 0. Si ha $\square$:

$$x\top_D y \leq x\top y \leq x \wedge y \ , \quad x \vee y \leq x \perp y \leq x \perp_D y$$

Come *corollario*, poiché il minimo è sempre $\leq$ del massimo, $x\top y \leq x \perp y$ anche se $\top$ e $\perp$ *non* sono duali.

Non si puó volere la botte piena e la moglie ubriaca.
Le due norme di Lukasiewicz verificano sia il principio di non contraddizione sia quello del terzo escluso. Un vantaggio non da poco, ma controbilanciato dal fatto che cadono due proprietà duali anch'esse "irrinunciabili", quelle di idempotenza $A = A\top A = A \perp A$. In realtà a qualcosa *bisogna* rinunciare. Supponiamo che la negazione di $1/2$ sia $1/2$, che valgano i due principi sacri della logica tradizionale e le due idempotenze (stiamo chiedendo poco, ma potremmo chiedere perfino di meno), e sottolineiamo ancora una volta che il v.l. (valore logico) di un'espressione logica dipende solo dai valori logici delle proposizioni coinvolte. Giungiamo purtroppo a due imbarazzanti assurdi!

$$1/2 \doteq \text{v.l.(P)} = \text{v.l.(P}\top\text{P)} = \text{v.l.(P}\top\neg\text{P)} = 0 \ , \quad \text{dualmente } \ 1/2 = 1$$

Universi infiniti*. Se gli universi sono infiniti le due norme duali standard non creano problemi, usando al caso estremi superiori e inferiori invece di massimi e di minimi. Se le norme non sono standard, e.g. sono norme alla Lukasiewicz, bisognerebbe invece ricorrere □ a serie o a integrali.

3.6 L'implicazione delle lingue naturali

Nella logica binaria tradizionale, che è la logica del vero e del falso (sì o no, 1 o 0, bianco o nero senza gradazioni di grigio) di *esprit de finesse* ce n'è pochino - e Aristotele, a onor del vero, sapeva bene che una proposizione come *Domani i greci vinceranno una battaglia navale*, non è nè vera nè falsa, è solo *possibile*, per cui la sua logica in effetti era ternaria, e niente affatto binaria. Un problema didattico che l'implicazione binaria pone ai discenti è che, come i docenti non si stancano di ripetere, dal falso si può dedurre tutto, sia il vero sia il falso, ed è quest'ultimo il punto dolente. Usando i codici 1 = vero e 0= falso, se il valore logico di P è 0 e il valore logico di Q è 1, la tabellina di calcolo restituisce 1 per il valore logico di $P \to Q$. Per tentare di convincere i discenti titubanti, un collega immaginativo fece appello al modo di dire triestino (e non solo): *se mia nona la gavessi le rodele la saria un tram*, che a dire il vero i triestini recitano anche nelle varianti, tutte corrette: *se mia nona la gavessi le rodele la fossi un tram*, *se mia nona la gaveria le rodele la saria un tram*, *se mia nona la gaveria le rodele la fossi un tram* (l'analogo turco *teyzemin taşağî olsaydî dayîm olurdu* è molto meno adatto all'uso in classe, perchè dice che se mia zia avesse immaginate cosa sarebbe mio zio). Ricapitolando: nel sillogismo binario, quello *hard* vincolato a non avere gradazioni di grigio, la fastidiosa presenza delle nonne e delle zie con attributi impropri provoca una giustificata reazione di perplessità, quasi che di implicazioni fosse serio parlare solo quando la premessa P è rigorosamente vera. Entrambe le perplessità sono molto importanti nella logica *fuzzy*, che al ragionamento di tipo verbale si ispira in maniera esplicita.

Visto che è dalle lingue naturali che traiamo ispirazione, prima di introdurre il sillogismo *fuzzy* vediamo di passare in rassegna come si comportano alcune lingue naturali, che dell'ipotetico linguaggio naturale "astratto" che sarebbe rinchiuso e codificato nei nostri collegamenti neuronali sono le esplicitazioni effettive con cui noi possiamo venire in contatto. Sarà una nuova occasione, dopo Bernoulli e Aristotele, di rendere omaggio agli antichi, ai grammatici cui si deve quell'analisi logica tradizionale che nei nuovi programmi di insegnamento è stata sostituita da analisi logiche molto più fini, ma forse solo più pretenziose, che talvolta provocano sconforto in chi di logica si intenda davvero - e che comunque i nostri ragazzi provvedono a obliterare rapidamente. I grammatici tradizionali distinguono due sole forme di implicazioni fondamentali, che chiamano il *periodo della realtà* e il *periodo dell'irrealtà*, proprio come succede nell'italiano parlato dai nostri ragazzi, che di congiuntivo non ne vogliono più sapere:

se vieni ti offro un caffè (periodo della realtà, affine a implicazioni matematiche del tipo: *se* 2n *non è divisibile per* 4*, allora l'intero* n *è dispari*)

se venivi ti offrivo un caffè (periodo dell'irrealtà, più vicino alle nonne e alle zie di cui sopra)

L'italiano con il congiuntivo offre invece (almeno) tre livelli:

se vieni ti offro un caffè (periodo della realtà)
se venissi ti offrirei un caffè
se fossi venuto ti avrei offerto un caffè (periodo dell'irrealtà)
La gradazione intermedia introduce un elemento di dubbio e/o di desiderio, ma in altri casi si sovrappone al periodo dell'irrealtà, quando l'antecedente è permanentemente falso: *se l'oro fosse a buon prezzo lo userei per farmici la vasca da bagno*. Ricchezza, ridondanza, confusione? Si devono evitare malintesi: le lingue sono (e devono essere) ridondanti e perfino ambigue, ne va del loro potere espressivo e dunque della loro forza comunicativa. Ma qui stiamo parlando di logica e di matematica: posto di fronte a una struttura reale, il matematico (o il logico) la sfronda di tutto ciò che è accessorio, in maniera da concentrarsi su ciò che è davvero essenziale e fondamentale. Insomma, ben fanno gli insegnanti a pretendere l'uso del congiuntivo, ma la ragioni di questa giusta pretesa non sono di tipo logico, prova ne sia una breve digressione in altre lingue, scelte in base ai gusti personali dell'autore. Il romeno, una lingua latina per certi aspetti arcaica (ha conservato tracce di declinazione), ma fortemente influenzata dalle lingue vicine, specie quelle slave, ha un'articolazione ternaria analoga a quella italiana (le traduzioni sgrammaticate ma letterali mirano a dare il senso di quel che succede):
dacă am timp merg la birou (se ho tempo vado in ufficio)
dacă aş avea timp aş merge la birou (se avrei tempo andrei in ufficio)
dacă aş fi avut timp aş fi mers la birou (se avrei avuto tempo sarei andato in ufficio
Solo binaria l'articolazione di base dello sloveno, che è una lingua slava come il russo o il polacco:
če prideš, ti plačam kavico (se vieni ti pago il caffè, periodo della realtà)
če bi prišel, bi ti plačal kavico (se verresti ti pagherei il caffè: periodo dell'irrealtà - il condizionale composto esiste, ma suona "pesante" nello sloveno parlato).
Per rimanere nella famiglia linguistica indoeuropea senza restare vincolati all'Europa, in persiano siamo molto vicini all'uso italiano colloquiale, anche se (ma non sempre) con un congiuntivo imprevisto:
agar zud biâyi otobus râ migiri (se tu venga presto, prendi l'autobus, realtà)
agar zudtar miâmadi otobus râ migerefti (se tu venivi prima, prendevi l'autobus)
Se qualcuno si fosse montato la testa dopo questi due esempi "primitivi" perchè soltanto binari, mi permetto di notare che l'articolazione binaria è stata sufficiente a Omar Khayyām per scrivere i suoi versi, come questi
gar âshegh o meykhâre be duzakh bâshad
fardâ bini behesht chon kaf e dast
(*se gli innamorati e i bevitori di vino andassero all'inferno, vedresti il paradiso come il palmo della mia mano* - ossia vuoto, senza le monete che il poeta avrebbe voluto vederci e che gli sarebbero servite per peccare; nel testo persiano Khayyām usa il periodo della realtà, congiuntivo nell'antecedente e indicativo nella conseguente, tanto evidente gli appare la concretezza dell'implicazione). Darò un solo esempio non indoeuropeo, quello dell'ebraico, che è una lingua semita come l'arabo, l'aramaico o l'egiziano antico. Rigorosamente binaria

l'articolazione, ma non definitela "primitiva" perchè le lingue semite sono state sufficienti per scrivere sia la Bibbia che il Corano!

im yesh lanu zman, anakhnu holkhim lakolnoa (se abbiamo tempo andiamo al cinema, periodo della realtà)

lu aba sheli haya rotze, hayinu nos'im la'amerika (se nostro padre era volente, eravamo andanti in America, periodo dell'irrealtà)

Dopo questo excursus nelle lingue naturali, torniamo all'implicazione *fuzzy*, un'implicazione artificiale che tuttavia all'implicazione naturale ha il dovere di esplicitamente ispirarsi.

3.7 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*:

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

Come si è visto, esiste tutta una teoria matematica delle operazioni logiche astratte, di cui quelle alla Zadeh o alla Luksiewicz sono casi speciali, e nella quale le congiunzioni e le disgiunzioni generalizzate vengono chiamate, rispettivamente, *t-norme* e *t-conorme*.

3.8 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 $A(x)$

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

[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.

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 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,

[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".

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 $h = 1/2$.

3.9 *Defuzzification*, focalizzazione, messa a fuoco

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.

3.10 Controllo *fuzzy*

Per il momento cfr. Bertoluzza.

3.11 Aritmetica *fuzzy*: l'approccio insiemistico.

Questa sezione riprende quanto detto sopra sulle *quantità sfocate*. L'approccio insiemistico all'aritmetica *fuzzy* è quello più tradizionale, ed è ancora presente nella letteratura. Ha però gravi debolezze, che verranno completamente rimosse nella sezione seguente, che descrive l'impostazione possibilistica.

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 $\{x, y : x \bullet y = z\}$; per capire quale sia il grado di appartenenza della *coppia* x, y all'insieme delle coppie $A \times B$ abbiamo bisogno della nozione di *prodotto cartesiano* (insieme delle coppie) di due insiemi sfocati, vedi sopra sezione??:

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

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, ma sarà del tutto superata da quanto diremo nella sezione seguente, dedicata all'approccio possibilistico. Segue un'altra lista di proprietà di cui la somma, il prodotto e la divisione fra quantità sfocate sfortunatamente *non* godono:

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

Inattesa dell'approccio possibilistico, il disastro è mitigato dal fatto che la proprietà distributiva vale se il primo termine è nitido: $r(A+B) = rA+rB$. Un *caveat*: 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. Dà (contro)esempi a sostegno delle mancate identità precedenti. Prendiamo il caso della mancata proprietà distributiva; si riesce a dimostrare soltanto:

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

ossia $A(B+C) \subseteq AB+AC$. Attenzione! La relazione precedente è un'inclusione fra insiemi sfocati: 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 può 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 $A^2 \subseteq A \times A$. Prima di imbarcarsi in calcoli, sarà meglio attendere le sezioni successive e il semplice quanto liberatorio *lemma di Montecatini*.

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" nelle applicazioni non ha grossi problemi' per il semplice fatto che è largamente triangolare e trapezoidale (lineare). Si potrebbe generalizzare, come si fa nella letteratura, introducendo "triangoli" e "trapezi" in cui i tratti ascendenti e discendenti non siano più vincolati alla linearità, e comunque imponendo che gli α-tagli siano intervalli nitidi chiusi, $\alpha \neq 0$. Che cosa debba essere un *numero fuzzy* in tutta generalità è un problema spinoso che troviamo piu adeguato trattare nelle sezioni successiove, dove potremo avvalerci di strumenti tecnici più comodi ma più forti.

3.12 Numeri *fuzzy*

Una quantità *fuzzy* X è completamente definita da ciò che cominceremo fin d'ora a chiamare la sua *funzione di distribuzione* (eventualmente sub-normale) $\mu_X(x) \doteq \Pi_X(x) \doteq X(x) : \mathbb{R} \to [0,1]$. Se la quantità sfocata $\underline{X}$ è n-dimensionale si ricorre a una funzione di distribuzione $\mu_{\underline{X}}(x_1 x_2 \dots x_n) \doteq \Pi_{\underline{X}}(x_1 x_2 \dots x_n) \doteq \underline{X}(x_1 x_2 \dots x_n) : \mathbb{R}^n \to [0,1]$. Rimane una perplessità: la definizione che abbiamo dato è talmente ampia da lasciar spazio a funzioni teratologiche, di quelle che piacciono agli analisti per costruire insidiosi controesempi che sfidano la nostra intuizione matematica. Circoscrivere le quantità *fuzzy* a una definizione sensata di cosa debbba essere un *numero fuzzy* è un problema spinoso che ha portato a soluzioni contrastanti nella letteratura. Offriremo tre punti di vista sempre più restrittivi fra cui il lettore potrà scegliere, anche se noi nel seguito ci atterremo alla seconda alternativa. Prima di farlo vogliamo tuttavia sottolineare che, se partiamo da una quantita sfocata n-dimensionale $\underline{X}$, ricordando quanto già detto sopra nella sezione ??, possiamo costruire n quantita sfocate *marginali* X_1, $X_2, \dots, Xn$ ponendo:

$$\Pi_{X_i}(x) \doteq \sup_{\underline{x}^{(i)} \in \mathbb{R}^{n-1}} \Pi_{\mathcal{X}}(x_1 x_2 \dots x_n) \ , \quad x_i = x$$

dove $\underline{x}^{(i)}$ è il vettore $(n-1)$-dimensionale ottenuto da $\underline{x}$ eliminando la sua componente i-esima x_i. Come succedeva sopra nel caso delle n proiezioni, le n distribuzioni *marginali* dei numeri sfocati X_i, $i = 1, 2, \dots, n$, *non* bastano per ricostruire la distribuzione congiunta n-dimensionale e uno dei modi possibili di procedere per ottenere una tale congiunta (in generale diversa da quella da cui si è partiti) è quello di "incollare" le marginali per *non-interattività* (termine pesante che deve ancora attendere la sua giustificazione fino alla sezione ??): $\Pi_{\underline{X}}(x_1 x_2 \dots x_n) \doteq \Pi_{X_1}(x_1) \wedge \Pi_{X_2}(x_2) \wedge \dots \wedge \Pi_{X_n}(x_n)$. Ed ecco le tre proposte, cui seguiranno le tre (contrastanti) motivazioni.

Proposta di definizione 1. Un numero fuzzy X è una qualunque quantita fuzzy normale.

Proposta di definizione 2. Un numero fuzzy X è una quantita fuzzy normale la cui funzione di distribuzione $\Pi(x)$ è vincolata ad essere *superiormente semi-*

continua e le cui α-sezioni $\{x;\ \Pi_X(x) \geq \alpha\}$ sono vincolate ad essere sottinsiemi (nitidi) *limitati* di $\mathbb{R}$ per $\alpha > 0$.

Proposta di definizione 3. Un numero fuzzy X è definito come nella proposta 2, imponendo inoltre quanto segue: tutte le α-sezioni sono insiemi *convessi* di $\mathbb{R}$ e inoltre l'equazione $\Pi_X(x) = 1$ ammette esattamente *una* soluzione (la funzione di distribuzione è *unimodale*).

Si noti che la convessità "quasi" implica l'unimodalità $\square$, se solo la 1-sezione non è un intervallo di ampiezza positiva, come succede nel caso dei *numeri fuzzy trapezoidali* ossia degli *intervalli fuzzy.* Nel caso di numeri *fuzzy n-dimensionali*, in particolare di coppie *fuzzy* XY, le tre proposte possono venir subito generalizzate $\square$. E passiamo alle tre motivazioni che potrebbero venir addotte da "*fuzzisiti*" rivali; il discorso aquisterà peso ulteriore nelle sezioni successive.

Motivazione della definizione 1. Nel calcolo della probabilità un numero aleatorio X (una variabile aleatoria) viene definito nel massimo della generalità possibile, compatibilmente con i soli vincoli tecnici imposti dalla teoria della misura. E' vero che il contenuto della definizione è troppo ampio e dà spazio a "mostri", ma è altrettanto vero che i teoremi significativi, imponendo per esempio l'esistenza di una varianzao finita, limitano di volta in volta la definizione (anche se la limitano il meno possibile). I nostri estremi inferiori e superiori non hanno i problemi delle somme e degli integrali, per cui la generalità, nel nostro caso, puo ben essere totale, rimanendo scontato che i teoremi significativi imporranno restrizioni anche pesanti. In effetti, potremmo perfino introdurre nel discorso numeri sfocati sub-normali, giungendo a una totale identificazione fra numeri e quantità.

Motivazione della definizione 2. Pensiamo ad esempio al caso bidimensionale XY. La definizione garantisce $\square$ che su ogni insieme chiuso, che sia limitato o meno, l'estremo superiore è anche un massimo: ora le rette orizzontali e verticali del piano reale $\mathbb{R}^2$ sono appunto chiuse, anche se illimitate. Ma allora la definizione 2 ci consente di marginalizzare (si riveda la def. ??) scrivendo nella formula un *max* al posto di un *sup*. Questo vantaggio tecnico si rivela spesso prezioso.

Motivazione della definizione 3. Le definizioni abituali presenti in letteratura sono del tipo 3, anche se con qualche piccola variante. L' idea è che un numero *fuzzy* dovrebbe rassomigliare il più possibile a un numero *crisp*, altrimenti il modello che stiamo usando rischia di essere stravagante. Rimangono esclusi in particolare gli intervalli *fuzzy* che modellano per l'appunto intervalli e non numeri.

Le sezioni seguenti ci permetteranno di approfondire la polemica con maggior cognizione di causa, ma ci si consenta subito un appunto all'ultima motivazione, per quanto popolare essa sia. Escludere numeri *fuzzy* multimodali significa rinunciare a numeri che modellano la seguente situazione: ci vediamo all'ora fuzzy X che, visti i miei turni di pausa dal lavoro, è grossomodo le 9 o indifferentemente le 11 di mattina. E perché non modellare il numero X con una funzione di distribuzione costituita da due triangoli disgiunti con vertici di ascissa 9 e 11? Un probabilista non avrebbe remore di questo tipo, e nel calcolo delle probabilità funzioni di densita di tipo trapezoidale non impediscono di parlare di numeri casuali, anche se il *numero* casuale in questo caso è *anche* un intervallo casuale: e perché no?

Finora abbiamo deplorevolmente sottaciuto quale debba essere la "forma" del *supporto* di X, $\{x : \Pi_X(x) \neq 0\}$, ma naturalmente il problema non si pò evitare. Una via d'uscita, indicate se si adotta la seconda definizione di un numero sfocato, potrebbe essere di ammettere unioni finite di insiemi connessi; anche se poco popolare in letteratura, siamo convinti che non andrebbero affatto esclusi supporti illimitati, lasciando spazio, per dire, a numeri sfocati *gaussiani*, definiti tramite una curva a campana di tipo gaussiano normalizzata a 1, ad esempio $\Pi(x) = \frac{1}{2\pi} \exp_e \frac{x^2}{2}$. In realtà abbiamo evaso anche il problema di quali sottinsiemi di $\mathbb{R}^n$ abbiano il diritto di chiamarsi *eventi*, evitando sottinsiemi teratologici così come si fa nel calcolo delle probabilità, dove ci si "limita" a sottinsiemi boreliani come definiti nella teoria della misura di Lebesgue. A dire il vero, la limitazione (assai generosa, peraltro) è dovuta a problemi tecnici di una teoria *additiva* com'è il calcolo della probabilità: la teoria delle possibilità è invece *maxitiva*, per cui non crea grane identificare sottinsiemi ed eventi, anche se poi solo di pochi e ben strutturati eventi avremo effettiva necessità.

3.13 Aritmetica fuzzy: l'approccio possibilistico.

Tornando all'impostazione insiemistica dell'aritmetica fuzzy (sez. ??), il lettore potrà notare che non rimane escluso che sia $X \times X \neq X^2$ (basta pensare a un numero *fuzzy* X di supporto parte positivo parte negativo $\Box$). L'approccio possibilistico, direttamente ispirato al calcolo delle probabilità e più nello specifico al calcolo delle distribuzioni di probabilità, elimina questa stranezza e permette di fondare l'aritmetica *fuzzy* su basi più solide da un punto di vista sia tecnico sia "filosofico". La nostra visione della situazione è che i numeri *fuzzy* dovrebbero ispirarsi ai numeri *casuali* (*variabili aleatorie*) più che ai numeri nitidi, come si è invece fatto fin dall'inizio dell'aritmetica *fuzzy*: è un punto di vista che si sta affermando sempre più nella letteratura della *fuzziness.* E in fondo il calcolo delle probabilità è nato nel '600, e dunque ha avuto ben tempo di assestarsi!

Per operare con un numero casuale (o variabile aletaoria unidimensionale; si confronti l'appendice alla fine del libro) si deve conoscere la sua *distribuzione di probabilità* proprio come per operare con una n-upla di numeri aleatori, per esempio quando li si vuole sommare, bisogna conoscerne la *distribuzione di probabilita congiunta.* Nell'aritmetica *random*, di solito chiamata *calcolo delle distribuzioni*, conoscere soltanto le distribuzioni *marginali* degli n addendi di per sé non basta. Di seguito adotteremo lo stesso punto di vista coi numeri sfocati; per liberarci da tecnicismi inutili ci serviremo del cosiddetto "approccio implicito" patrocinato dal matematico di Cambridge H.R. Pitt (1914-2005), in cui, senza dire "esplicitamente" che cosa essa *sia*, la variabile aleatoria X viene "implicitamente" definita assegando solo la sua distribuzione di probabilità P_X su $\mathbb{R}$ (su $\mathbb{R}^n$ nel caso di n-uple aleatorie) .[7] Esattamente allo stesso modo, implicitamente ma *esaurientemente*, definiremo i numeri sfocati tramite le loro distribuzioni: l'analogo di P_X nella letteratura *fuzzy* è un insieme sfocato normale (si riveda la sezione??), che talvolta viene chiamato la *funzione caratterizzante* di X e talvolta, come faremo noi, la sua *distribuzione di possibilità* Π_X: è una questione di nomi, potremmo anche parlare di insiemi, ma troviamo suggestiva la contrapposizione:

numero aleatorio verso *numero sfocato*
sua distribuzione di probabilità verso *sua distribuzione di possibilità*

Sulle distribuzioni di possibilità

$$\mathrm{Poss}\{\mathrm{X} \in \mathrm{E}\} \doteq \Pi_{\mathrm{X}}(\mathrm{E}) \ , \quad \mathrm{Poss}\{\mathrm{XY} \in \mathrm{F}\} = \Pi_{\mathrm{XY}}(\mathrm{F}) \ , \quad \mathrm{E} \subseteq \mathbb{R}, \ \ \mathrm{F} \subseteq \mathbb{R}^2$$

si veda anche l'Appendice alla fine della sezione: si noti che senza pericolo di confusione, stiamo usando lo stesso simbolo Π sia per la funzione $\Pi(x): \mathbb{R} \to [0,1]$ sia per la sua estensione $\Pi(E): 2^{\mathbb{R}} \to [0,1]$ a sottinsiemi o eventi $E \subseteq \mathbb{R}$.

Ribadiamo che, proprio come nel caso probabilistico, per sommare o moltiplicare due numeri sfocati X e Y dobbiamo conoscerne la distribuzione di possibilità congiunta Π_{XY}, non bastano le due marginali Π_X e Π_Y, che dalla congiunta si derivano calcolando il massimo (al caso l'estremo superiore) sulle rette vericali e orizzontali, si riveda la sezione ?? e la nota ?? Data che sia una generica operazione binaria $\circ : \mathbb{R}^2 \to \mathbb{R}$ fra due operandi sfocati X e Y, ad esempio una somma o un prodotto, il risultato $Z = X \circ Y$ è anch'esso un numero *fuzzy*, per cui, per controllarlo esaurientemente, occorre e basta conoscere la sua distribuzione di possibilità Π_Z. Stante l'equivalenza dei due eventi $F \doteq \{Z = z\}$ e $\{XY \in C_z\}$ dove $C_z \doteq \{(x,y) : x \circ y = z\}$, (essi si implicano l'un l'altro), e stante la "maxitività" delle possibilità, poniamo com'è naturale:

$$\mathrm{Poss}\{Z = z\} \doteq \Pi_Z(z) \doteq \Pi_Z(F) = \Pi_{XY}\{XY \in C_z\} \doteq \max_{x,y \in C_z} \Pi_{XY}(x,y)$$

[7] Nell'approccio "esplicito" di Kolmogorov, le variabili aleatorie sono definite grazie a funzioni boreliane, un concetto della teoria della misura relativamente avanzato; anche se in linea di principio l'approccio di Kolmogorov è di solito quello ufficialmente seguito, in pratica ci si comporta spessissimo alla Pitt, ad esempio quando si assegna una variabile gaussiana senza sentire nessun bisogno di esplicitare qualsivoglia funzione borelina. Si conviene che una notazione suggestiva come $\mathrm{Prob}\{X \in E\}$ voglia semplicemente dire $P_X(E)$ per ogni *evento* E, vale a dire per ogni insieme boreliano $E \subseteq \mathbb{R}$. Formalmente, è facile dimostrare che le due impostazioni sono del tutto equivalenti, ma conviene rifarsi a Kolmogorov quando si ha a che fare con intricati processi stocastici. Le acque in cui qui noi ci muoviamo sono fortunatamente molto pú tranquille!

Va sottolineato che invece di ricorrere al principio di estensione di Zadeh ci stiamo ispirando ad argomenti tipici del calcolo delle distributioni probabilistico. Per poter scrivere *max* invece di *sup*, oltre che ricorrere alla def. ?? di un numero sfocato converrà anche ipotizzare:

tutte le controimmagini $C_z \doteq \{x, y : x \circ y = z\}$ *sono sottinsiemi* chiusi *di* $\mathbb{R}^2$

Prima di passare al liberatorio lemma di Montecatini, elenchiamo tre notevoli distribuzioni congiunte che "incollano" due marginali assegnate Π_X e Π_Y; si rammenti che la non-interattività, per la quale usiamo le notazione $\Pi_{X\perp Y}$, è considerata l'analogo possibilistico dell'independenza probabilistica.

Non-interattività: $\Pi_{X\perp Y}(x, y) \doteq \Pi_X(x) \wedge \Pi_Y(y)$

Interattività drastica: Per x^* e y^* tali che $\Pi_X(x^*) = \Pi_Y(y^*) = 1$ si ponga:
$\Pi_{XY}(x^*, y) = \Pi_Y(y)$, $\Pi_{XY}(x, y^*) = \Pi_X(x)$, se no $\Pi_{XY}(x, y) = 0$

Uguaglianza deterministica: Assumendo che X and Y siano *equidistribuite*, vale a dire $\Pi_X = \Pi_Y$), si ponga:
$\Pi_{XY}(x, x) = \Pi_X(x) = \Pi_Y(x)$, se no $\Pi_{XY}(x, y) = 0$

Sottolineiamo con forza che due numeri fuzzi *distinti* $X \neq Y$ potrebbero ben avere la *stessa identica* distribuzione di possibilità $\Pi_X = \Pi_Y$, ma per operare sulla coppia XY *dobbiamo* specificare come i due termini coinvolti interagiscano fra di loro, il che può andare dalla non-interattività fino alla uguaglianza deterministica, e *solo in quest'ultimo caso* si ha il diritto di scrivere $X = Y$, per cui anche, ad esempio, $X \times Y = X \times X \doteq X^2$.

Il lemma che segue diede luogo a un'animata discussione a un incontro svoltosi a Montecatini. *Identità* che coinvolgono k numeri reali sono ad esempio: $x + x = 2x$, $x^2 = x \times x$, $x + y = y + x$, $x(y + z) = xy + xz$, $\log_2 xy = \log_2 x + \log_2 z$ per $= 1, 1, 2, 3, 2$ rispettivamente; va da sé che l'ultima richiede $x, y > 0$. Nel seguito quando ad esempio scriviamo $\log X$ oppure $\log XY$ sottintendiamo che i supporti coivolti debbano essere positivi: $\Pi_X(x) > 0 \Rightarrow x > 0$, $\Pi_{XY}(x, y) > 0 \Rightarrow xy > 0$. Nel lemma $f(\underline{x}) = g(\underline{x})$ indica una generica identità che coinvolge la k-upla numerica $\underline{x} \doteq x_1 x_2 \ldots x_k$. Per ribadire, quando scriviamo $f(\underline{X}) = g(\underline{X})$ dove la $\underline{X}$ è una k-upla *fuzzy*, sottintendiamo che il supporto della $\underline{X}$ sia vincolato ad essere incluso nel dominio della $f(\underline{x})$, ossia là dove la $f(\underline{x})$ è ben definita. Rammentiamo che il segno di $=$ fra due numeri sfocati indica un'uguaglianza deterministica e non la mera equidistribuzione.

Il lemma di Montecatini.
$f(\underline{x}) = g(\underline{x})$ se e solo se $f(\underline{X}) = g(\underline{X})$
Dimostrazione. L'implicazione da destra a sinistra è ovvia, perché i numeri *crisp* o nitidi sono un caso particolare di quelli *fuzzy* (in altre parole: nulla vieta che i numeri sfocati a destra del segno di $=$ siano in realtà nitidi). Per l'implicazione

inversa cominciamo con il dimostrare che $Z_f \doteq f(\underline{X})$ e $Z_g \doteq g(\underline{X})$ sono equidistribuiti $\Box$ (entrambi sono calcolati mediante la stessa e identica formula). Sono anche deterministicamente uguali: basta calcolare $\Box$ la distribuzione di possibilità congiunta della coppia (Z_f, Z_g) che è funzione deterministica della k-upla $\underline{X}$: $(Z_f, Z_g) = h(\underline{X})$. Se $\underline{x} \doteq x_1 x_2 \ldots x_k$ e $\underline{y} \doteq y_1 y_2 \ldots y_k$ non coincidono componente per componente la distribuzione congiunta bidimensionale di $[Z_f, Z_g]$ è nulla in $(u, v) \doteq [f(\underline{x}), g(\underline{y})]$: $[Z_f, Z_g](u, v) = 0$.

Come la mettiamo con le uguaglianze mancate della sezione ??, ad esempio con la $X^2 \neq X \times X$ o con la $X(Y + Z) \neq XY + XZ$, che sono in (apparente) contraddizione col lemma di Montecatini? Semplice quanto deplorevole: l'impostazione più tradizionale, quella che abbiamo chiamato insiemistica, non distingue fra deterministica uguaglianza ed equidistribuzione, e le due mancate uguaglianze andrebbero invece scritte: $\Pi_X = \Pi_U$ non implica né $\Pi_{X^2} = \Pi_{X \times U}$ né $\Pi_{X(Y+Z)} = \Pi_{XY+UZ}$. Scritte così non turbano piu nessuno: neanche nel calcolo delle probabilità (dove vige l'analogo del lemma di Montecatini) l'equidistribuzione di due variabili aleatorie X e U che potrebbero ben essere fortemente dipendenti implica che X^2 e $X \times U$ siano la *medesima* variabile aleatoria, e non implica neppure che abbiano almeno la stessa distribuzione. Altro è il discorso se X e U sono semplicemente due nomi diversi per lo stesso oggetto (altro è il caso dell'uguaglianza deterministica): allora sì che, sia in probabilità sia nell'artitmetica *fuzzy*, si ha quello che tutti si aspettano, ossia $X^2 = X \times U \doteq X \times X$. Una domanda $\Box$: è un caso che in tutte le uguaglianze mancate della sezione?? la stessa lettera maiuscola compaia almeno due volte in un unico membro delle (pretese) disuguaglianze?

3.14 Irrilevanza dell'interattività

L'aritmetica sfocate tradizionale, quella insiemistica, conosce di fatto solo due modi in cui due distribuzioni marginali Π_X e Π_Y possono venir "incollate" a formare una congiunta Π_{XY}, ossia la non-interattività (confronta le formule ?? e ??) e l'uguaglianza deterministica, quest'ultima possibile solo se si ipotizza che le due marginali coincidano, $\Pi_X = \Pi_Y$. Purtroppo, in quest'ultimo caso, non si sa distinguere fra la non-interattività di due numeri sfocati equidistribuiti e la loro uguaglianza deterministica, quando i due numeri sono in realtà la stessa cosa, mentre due numeri, pur equidistribuiti, ma non-interattivi sono tutt'altro che la stessa cosa, visto che "non si conoscono". Ora il problema dell'irrilevanza cui è dedicata la sezione si occupa della seguente situazione: è data una famiglia di distribuzioni bidimensionali congiunte tutte con la stessa marginale "orizzontale" Π_X e la stessa marginale "verticale" Π_Y; se si vuole, di due numeri sfocati X e Y si conosce il comportamento marginale ma non si sa bene il modo in cui interagiscono. Può succedere che il risultato di un'operazione $Z = X \circ Y$ sia lo stesso indipendentemente da quale interazione leghi X e Y all'interno della famiglia di congiunte prescelta? Il problema è ampio, ma noi ci limiteremo all'unica situazione interessante dal punto di vista dell'aritmetica sfocata tradizionale, vale a dire la situazione in cui le due marginali coincidano, $\Pi_X = \Pi_Y$, e la famiglia delle congiunte si riduca a due soli elementi, la non-interattività $\Pi_{X\perp Y}$ e la deterministica uguaglianza $\Pi_{X=Y}$.

Il teorema che segue copre molti casi di numeri fuzzy usati in pratica, a partire da quelli triangolari. La notazione $\simeq$ indica mera equidistribuzione e non uguaglianza deterministica $=$; dicendo di un'operazione $\circ$ che *conserva l'ordine*, ad esempio l'ordine non decrescente, ossia debolmente crescente, intendiamo dire che se $x < y$ allora si ha $x \circ u \leq y \circ u$ e $u \circ x \leq u \circ y$ per ogni possibile u.

Teorema. Siano X and Y equidistribuiti e sia $\Pi_X(x) = \Pi_Y(x)$ una funzione concava (*convex-cap*, convessa a cappello) sul suo supporto *connesso*. Sia $\circ$ un'operazione che conserva l'ordine tale che $f(u) = u \circ u \doteq u^{(2)}$sia una funzione continua del suo argomento. Si ha allora $X \circ Y \simeq X \circ X \doteq X^{(2)}$.
Dimostrazione. Supponendo ad esempio che l'operazione $\circ$ sia non-decrescente, non-decrescente è anche la funzione $f(u)$. Se $x < y$, si ha $x^{(2)} \leq x \circ y \leq y^{(2)}$, per cui, stante la continuità di f, deve esserci un valore $u = \alpha x + (1 - \alpha)y$ intermedio fra x e y $(0 \leq \alpha \leq 1)$ dove $u^{(2)} = x \circ y$; ma allora par concavità $\Pi_X(u) \geq \alpha\Pi_X(x) + (1 - \alpha)\Pi_X(y) \geq \Pi_X(x) \wedge \Pi_X(y) \doteq \Pi_{X\perp Y}(x, y)$. QED

Il teorema si applica sia alla somma $(\circ = +)$ sia al prodotto $(\circ = \times)$, in quest'ultimo caso di fattori con supporto comune positivo (non negativo): la prima affermazione è ovvia, e in quanto alla seconda si pensi $\Box$ al *prodotto positivo* che coincide col prodotto normale $x \times y$ se entrambi i fattori sono positivi, se no vale 0. Ora, se il supporto comune è non-negativi fra i due prodotti non c'è differenza alcuna. Rimanendo al prodotto, se il supporto è non-positivo, le cose vanno bene comunque $\Box$: si usi l'identità $X \times Y = (-X) \times (-Y)$. Si era invece già fatto notare nella sezione?? che se il supporto è misto (esistono $x_1 < 0$ e $x_2 > 0$ di possibilità positiva), l'irrilevanza *non* vale per la moltiplicazione, visto che $z \doteq x_1 \times x_2 < 0$ ha possibilità positiva solo con la non-interattività, mentre se c'è uguaglianza deterministica $X \times X \doteq X^2$ è un quadrato, e il suo supporto è ovviamente e interamente non-negativo.

Siano $X \simeq Y$ triangolari su $[a, c]$ con vertice in b, $0 \leq a < b < c$. Si ha irrelevanza dell'uguaglianza deterministica rispetto alla non-interattività sia per la somma che per il prodotto, ma *non* si ha irrelevanza *tout court*. Si prenda ad esempio una distribuzione *drastica* come definita in??: in tal caso $\Box$ il supporto di $X+Y$ terminerebbe a $b + c$ e non a $2c$, mentre il supporto di $X \times Y$ terminerebbe a $b \times c$ e non a c^2.

Esercizio. Data la distribuzione triangolare (a, b, c) comune agli operandi X e Y si calcoli esplicitamente la distribuzione di somma e prodotto, distinguendo i tre casi in cui X e Y siano non-interattivi, deterministicamente uguali e drasticamente interagenti. In quanto al prodotto, il lettore impaziente preferirà limitare i calcoli al caso $a \geq 0$, perché allora il risultato per la non-interattività e per l'uguaglianza deterministica coincidono.

Appendice sulle possibilità.
In quest'appendice troviamo conveniente generalizzare i numeri *fuzzy* a *quantitià fuzzy* o perfino ad *attributi fuzzy* X non richiedendo nulla alla distribuzione di possibilità Π_X sull'"universo" $\mathcal{U}$, universo che potrebbe non essere neppure numerico; poniamo $\text{Poss}\{X \in E\} \doteq \Pi_X(E)$ per ogni evento E (per ogni sottinsieme nitido $E \subset= \mathcal{U}$). Sia $f : \mathcal{U} \to \mathcal{V}$ una funzione deterministica da $\mathcal{U}$ a $\mathcal{V}$, e sia X un'attributo *fuzzy* su $\mathcal{U}$. Se l'attributo *fuzzy* Y su $\mathcal{V}$ è definito da $Y \doteq f(X)$, calcoliamone la distribuzione Π_Y. Si scelga $y \in \mathcal{V}$: ora, i due eventi $E \doteq \{x : f(x) = y\} \subseteq \mathcal{U}$ e $\{y\} \subseteq \mathcal{V}$ si implicano l'un l'altro, per cui devono avere la stessa possibilità. Ne viene subito $\text{Poss}\{Y = y\} = \text{Poss}\{X \in E\} = \sup_{x:\, f(x)=y} \Pi_X(x)$. Se $\mathcal{U}$ è il quadrato cartesiano $\mathbb{R}^2$, $\mathcal{V} = \mathbb{R}$ and $f = \circ$, si ritrova (??) sopra nella sezione??.
Un poset. Dato $\mathcal{U}$, si trova subito un *poset* (*p*artially *o*rdered *set*, insieme parzialmente ordinato) sullo spazio di tutte le distribuzioni di possibilità su $\mathcal{U}$ ponendo $\Pi_1 \leq \Pi_2$ sse $\Pi_1(x) \leq \Pi_2(x)$ $\forall x \in \mathcal{U}$, e dunque $\forall E \subseteq \mathcal{U}$. Nella famiglia $\mathcal{F}$ di tutte le distribuzioni congiunte con marginali prefissati Π_X e Π_Y, la non-interattività $\Pi_{X \perp Y}$ è l'unico massimo nell'ordinamento del *poset*, mentre la distribuzione drastica, oppure, ipotizzando l' equidistribuzione $\Pi_X = \Pi_Y$, l'uguaglianza deterministica sono elementi minimali $\Box$. Si dimostra $\Box$:
Proposizione
If $\Pi_1 \leq \Pi_2 \leq \Pi_3$ and $\Pi_{f(\Pi_1)} = \Pi_{f(\Pi_3)}$ then $\Pi_{f(\Pi_1)} = \Pi_{f(\Pi_2)} = \Pi_{f(\Pi_3)}$
Non interattività e indipendenza. Si ritiene che la non interattività sia l' analogo possibilistico "naturale" dell' independenza probabilistica. Vediamo di argomentarlo. Combinare (incollare) due marginali $\Pi_1(x)$ e $\Pi_2(y)$ per non-interattività significa assegnare a tutte le coppie $(x, y) \in \mathcal{U} \times \mathcal{V}$ una possibilità congiunta $\Pi(x, y)$ che è la *massima* possibile. Se il nostro stato di conoscenza si riduce ai due marginali, non possiamo comprometterci a dir nulla della possibilità di (x, y) tranne ciò che davvero sappiamo, vale a dire che questa possibilità al più è $\Pi_1(x) \wedge \Pi_1(x)$. Abbassare questo valore (ma a qual altro valore?) in assenza di ulteriori informazioni sarebbe *unfair*.

Chapter 4

Misure d'informazione

4.1 L'entropia di Shannon

Il concetto di *entropia* ha un fascino irresistibile: esso interviene nei contesti più disparati, termodinamica, ecologia, informazione, spesso con tinte catastrofiche (morte dell'universo, caos informazionale). Non è neppure chiaro *quante* siano le entropie, se una o due (entropia termodinamica, entropia informazionale). La protagonista di questo libro è l'entropia informazionale, di Shannon. Diamo per scontato che il lettore sappia, almeno per sommi capi, che cos'è la probabilità, cui ad ogni buon conto abbiamo dedicato un'appendice.

Sia pure in sordina (in sordina nel senso che la *semantica* dell'entropia, il suo significato, ancora non si vedrà), entriamo subito nel vivo, e introduciamo l'*entropia di Shannon.*

L'*universo del* nostro *discorso* è molto semplicemente un insieme finito $\mathcal{A} = \{a_1, a_2, \ldots, a_K\}$ di K oggetti, che spesso troveremo suggestivo chiamare *lettere* dell'*alfabeto*$\mathcal{A}lA$; salvo specifica contraria $K \geq 2$. Ogni lettera a_i ha la probabilità $p_i = \text{Prob}\{a_i\}$ di venir scelta; $\forall i\ p_i \geq 0$, $\sum_i p_i = 1$, per cui $p_i \leq 1$. (Specifiche come la $\forall i$ verranno spesso sottintese quando sono ovvie nel contesto; sommatorie tacite si intendono estese a tutti i valori dell'indice; qui i va da 1 a K.)

Variare il *vettore di probabilità* P, che è un vettore reale a K componenti, significa variare la legge probabilistica che regola la scelta aleatoria delle lettere di $\mathcal{A}$. L'insieme di *tutti* i vettori di probabilità (di tutte le leggi aleatorie possibili su $\mathcal{A}$) è il dominio di definizione dell'entropia, il suo *ambiente.* Si tratta di un sottospazio *compatto* (ossia chiuso e limitato) e *convesso* di $\Re^K$ ($\Box$!). *Figura* per $K = 2, 3$.

Sia allora P un vettore di probabilità su $\mathcal{A}$; si pone:

$$H(P) \doteq -\sum_i p_i \log_2 p_i \equiv \sum_i p_i \log_2 \frac{1}{p_i}$$

con la convenzione di continuità $0 \log_2 0 = 0$ (e infatti $\lim_{x \to 0^+} x \log x = 0$). D'ora in poi ometteremo l'indicazione della base 2 dei logaritmi (sulla scelta della base ritorneremo fra poco; certo si è che in un contesto cosıspesso binario i logaritmi binari si impongono da sé). In particolare, per $K = 2$, $\mathcal{A} = \{0, 1\}$, $P = (1 - x, x)$, $0 \le x \le 1$:

$$H(P) = h(x) = -x \log_2 x - (1 - x) \log_2 (1 - x)$$

La $h(x)$ si chiama *funzione entropica*.

Figura. Grafico di $h(x)$. La funzione $h(x)$ è simmetrica, concava, ha pendenza infinita in 0 e 1, dove dunque non è derivabile. Inoltre: $0 \le h(x) \le 1$; i minimi corrispondono alle due probabilità *deterministiche*, o *degeneri*, il massimo alla probabilità *uniforme*.

In generale:

$$0 \le H(P) \le \log K$$

con $H(P) = 0$ sse P è deterministica, $H(P) = \log K$ sse P è uniforme (le dimostrazioni sono relegate agli esercizi in calce); $H(P)$ è una funzione concava di P (la dimostrazione è rinviata a quando parleremo di *entropie condizionate*).

Figura. Se $K = 3$ l'entropia ricorda un *wigwam* concavo che i pellerossa hanno attaccato a terra nei tre vertici del triangolo che fa da pavimento, e che ha tre porte verticali spalancate a forma di funzione entropica, ciøche fa un gran giro d'aria.

> Entropia e funzione entropica sono nomi impegnativi, di sapore termodinamico, che a suo tempo cercheremo di giustificare. Consigliamo fin d'ora, per i vantaggi euristici che comporta, di vedere l'entropia di P come una *misura dell'incertezza* associata all'esito dell'esperimento aleatorio X di legge probabilistica P che lo sperimentatore sta per effettuare. Il lettore apprezzerà il fatto che sotto questa luce le proprietà già dimostrate acquistano un significato perspicuo. In particolare è giusto che l'incertezza sia nulla in corrispondenza a situazioni in cui il risultato dell'esperimento è noto deterministicamente già a priori, e sia massima quando gli esiti a priori possibili sono equiprobabili, e non si sa davvero "che pesci pigliare".
>
> Claude Shannon pubblicò nel 1948 un articolo, *A mathematical theory of communication*, che fonda la *teoria dell'informazione*, e prefigura la teoria dei codici. La lettera H (acca) era in origine un'eta greca maiuscola, prima lettera di $\eta\nu\tau\rho\omega\pi\iota\sigma$.

Esercizio. Se l'esperimento consiste nel lancio di una moneta equa, $H(P) = 1$; se invece consiste nel lancio di n monete eque $H(P) = n$ (a priori ci sono 2^n esiti possibili ed equiprobabili).

> L'esercizio precedente dà una chiara idea di quale sia l'ordine di grandezza della nostra *unità di misura* per l'incertezza: essa si chiama *bit* (da *binary unit*). Se nella formula dell'entropia si usano i logaritmi in base $\beta > 1$, si dice che l'entropia è misurata in β-it, in particolare in *nat* (= *natural units*) nel caso dei logaritmi naturali.

Esercizio. Se $H(P)$ e $H_\beta(P)$ sono l'entropia di P in bit e in β-it, $H(P) = \alpha H_\beta(P)$, con α costante positiva indipendente da P (calcolala!). E' la tipica legge di trasformazione che si ha quando si cambia l'unità di misura.

Esercizio. Dimostra la disuguaglianza $H(P) \geq 0$, e stabilisci le condizioni di uguaglianza. Cenno: una somma di K addendi non negativi è nulla sse lo sono tutti i K addendi.

Esercizio-lemma 1. Dimostra la disuguaglianza $\ln x \leq x-1$, con uguaglianza sse $x = 1$ (il logaritmo naturale è una funzione strettamente concava che sta sotto la sua tangente nel punto di ascissa 1 e la tocca solo nel punto di tangenza).

Esercizio-lemma 2. Dimostra la disuguaglianza $\ln x \geq 1 - \frac{1}{x}$, con uguaglianza sse $x = 1$ (nel lemma precedente metti $\frac{1}{x}$ al posto di x).

Esercizio. Dimostra la disuguaglianza $H(P) \leq \log K$, e stabilisci le condizioni di uguaglianza. Cenno: pensando dapprima all'entropia misurata in nat, applica K volte il lemma 2 ai K fattori di $\ln K - H_e(P) = \sum p_i \ln(Kp_i)$.

Complementi stellati

In questa sezione anticipiamo la *formula di diramazione* dell'entropia. Essa consente di ridurre il calcolo dell'entropia di un vettore a K componenti al calcolo delle entropie di due vettori con meno componenti.

Se A è un sottinsieme proprio di $\mathcal{A}$ di probabilità P positiva (almeno una lettera a di A ha probabilità positiva; $A \subset \mathcal{A}$), possiamo costruire un nuovo vettore di probabilità P_A sull'alfabeto *parziale* A ponendo $P_A(a) = \frac{P(a)}{P(A)}$ (e se si vuole $P(a) = 0$ per $a \notin A$).

> Il lettore smaliziato avrà riconosciuto in P_A una *probabilità condizionata*, quando si ipotizza il verificarsi dell'evento A. L'agnizione è essenziale, ma per il momento i nostri interessi vanno solo all'uso delle tabelle. (Alle probabilità condizionate viene dedicata qualche parola nell'appendice.)

Oltre all'alfabeto parziale, ci serviremo dell'alfabeto *ingrossato* $\mathcal{A}^*$ di $K-|A|+1$ lettere, in cui le $|A|$ lettere di A vengono considerate "indistinguibili" (vengono "fuse" in un'unica "letterona", la A appunto); anche la probabilità P puøvenir ingrossata, ponendo $P^*(a) = P(a)$ per $a \notin A$, $P^*(A) = P(A) = \sum_{a \in A} P(a)$.

Formula di diramazione: $H(P) = H(P^*)+P(A)H(P_A)$ La formula è un'immediata

conseguenza delle proprietà elementari dei logaritmi (□ !) e consente di calcolare l'entropia di un vettore a K componenti usando $K-1$ volte la tabella 2. Il lettore rifletta alla sua interpretazione euristica in termini di incertezze.

Tabella 1. Tabula i valori di $-x \log_2 x$ da $x = 0$ a $x = 1$ con passo Δ.

Tabella 2. Tabula i valori di $h(x)$ da $x = 0$ a $x = \frac{1}{2}$ con passo Δ.

L'entropia $H(P)$ è una funzione continua di P definita su uno spazio compatto, e dunque è *uniformemente* continua. A soddisfazione degli analisti numerici, l'uniformità viene precisata e concretizzata nel risultato che segue; la limitazione superiore a destra della disuguaglianza non dipende né da P né da Q e tende a 0 insieme con la distanza fra P e Q. (La dimostrazione è "tecnica" e noiosa, per cui il lettore puøcomprare il risultato "a scatola chiusa".)

Faremo riferimento alla distanza di Manhattan $d_M(P,Q) = \sum_i |p_i - q_i|$ (a Trieste diremmo distanza del Borgo Teresiano; le ragioni del nome sono chiare non appena si faccia uno schizzo per $K = 2$). In rapporto alla "normale" distanza euclidea fra vettori di probabilità $d_E(P,Q) = \sqrt{\sum_i (p_i - q_i)^2}$ si ha $d_E \le d_M \le K d_E$, per cui le due *metriche* $d_E(P,Q)$ e $d_M(P,Q)$ definiscono la stessa *topologia* (□ ! cenni: da una parte $d_E^2 \le d_M^2$, dall'altra $d_E \ge \max_i |p_i - q_i| \ge \frac{d_M}{K}$).

Disuguaglianza: uniforme continuità dell'entropia. Se $d_M(P,Q) \le \frac{1}{2}$, si ha

$|H(P) - H(Q)| \le -d_M(P,Q) \log \frac{d_M(P,Q)}{K}$

Dimostrazione. Partiamo da una premessa analitica piuttosto noiosa che il lettore di buon gusto scorrerà di malanimo. Anzi, prima della premessa, per poi non interromperci, facciamo notare che $r \le s \le t \Rightarrow |s| \le \max[-r, t]$ (□ ! distingui i due casi $s \ge 0$, $s \le 0$). Se ora $f(x)$ è una funzione concava del numero reale x, e se $[a, a+\Delta]$ e $[x, x+\Delta]$, con $a \le x$ e $\Delta \ge 0$, sono due intervalli dell'insieme (intervallo esso stesso, semiretta o retta) su cui la funzione f è definita, si ha $f(x+\Delta) - f(x) \le f(a+\Delta) - f(a)$ (essenzialmente questa è la definizione di funzione concava). Passando a *tre* intervalli di ampiezza Δ e punti di partenza $a \le x \le b$, si trova $f(b+\Delta) - f(b) \le f(x+\Delta) - f(x) \le f(a+\Delta) - f(a)$. Dunque, per il valore assoluto $|f(x+\Delta) - f(x)|$:

$$|f(x+\Delta) - f(x)| \le \max[f(a+\Delta) - f(a), f(b) - f(b+\Delta)]$$

Applichiamo la disuguaglianza appena ottenuta alla funzione concava $f(x) = -x \log x$ con $a = 0$, $b + \Delta = 1$ $\big($al solito $f(0) \doteq 0$ per continuità; calcola la derivata seconda per dimostrare la concavità di $f(x)\big)$. Notato che $f(0) = f(1) = 0$, la maggiorazione precedente si scrive:

$$|f(x+\Delta) - f(x)| \leq \max[f(\Delta), f(1-\Delta)]$$

La lotta per il massimo è vinta da $f(\Delta)$ sull'intervallo $[0, \frac{1}{2}]$, da $f(1-\Delta)$ sull'intervallo $[\frac{1}{2}, 1]$ ($\square$!; confronta i grafici delle funzioni $f(x)$ e $f(1-x)$). Dunque, e ciøconclude la parte noiosa:

$$|f(x+\Delta) - f(x)| \leq f(\Delta)\ , \quad 0 \leq \Delta \leq \frac{1}{2}$$

Posto $\delta_i = |p_i - q_i|$, useremo questa disuguaglianza K volte; $\sum_i \delta_i \doteq d_M(P,Q) \equiv d_M \leq \frac{1}{2}$ garantisce $\delta_i \leq \frac{1}{2}$. Si ha: $|H(P) - H(Q)| \leq \sum_i |f(p_i) - f(q_i)|$, e dunque:

$$|H(P) - H(Q)| \leq \sum_i f(\delta_i) \equiv -\sum_i \delta_i \log \delta_i$$

Il membro di destra si puøscrivere in maniera elegante, e in apparenza stravagante, dopo aver introdotto il vettore di probabilità (*sic* !) $R \doteq (\frac{\delta_1}{d_M}, \frac{\delta_2}{d_M}, \ldots, \frac{\delta_K}{d_M})$:

$$|H(P) - H(Q)| \leq d_M\big(H(R) - \log d_M\big) \leq d_M(\log K - \log d_M) \equiv -d_M \log \frac{d_M}{K}$$

L'ultima disuguaglianza è la limitazione superiore naturale di qualunque entropia, anche di un entropia "artificiosa" come $H(R)$. (A esser pignoli, nella dimostrazione dobbiamo supporre $d_M \neq 0$, ma se $d_M = 0$ non c'è nulla da dimostrare, visto che in questo caso $P = Q$.)

Un semplice studio di funzione e la $d_M \leq K d_E$ danno subito i due corollari:

Corollario 1. Se $d_M(P,Q) \leq \epsilon \leq \frac{1}{e} \approx 0,37$, si ha $|H(P) - H(Q)| \leq -\epsilon \log \frac{\epsilon}{K}$

Corollario 2 (distanza euclidea). Se $d_E(P,Q) \leq \tau \leq \frac{1}{Ke}$, si ha
$|H(P) - H(Q)| \leq -K\tau \log \tau$

4.2 Misure d' ignoranza

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 $m(\mathcal{A}lX)$. Si ha $N(m) = m(\mathcal{A}lX)\log|\mathcal{A}lX|$, $D(m) = -\sum_i p_i \log\big(p_i + m(\mathcal{A}lX)\big)$, $C(m) = -\sum_i p_i \log p_i$.

Purtroppo il carattere piuttosto ad hoc delle ultime due proposte, $D(m)$ e $C(m)$, è ormai ammesso, ciøche 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 ($\Box$!); 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 ($\Box$!). 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. $\Box$

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 ($\Box$!): $D(m) \leq C(m)$

$D(m) = C(m)$ 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 perønon 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 $2^K - 1$ 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 calce 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!

Complemento stellato. 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 puøvenir 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 $C(m)$.

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 probabilitị $\mathcal{P}$, e lo si puøassumere come un misura dell'incertezza globale della probabilitıntervallare, o, al caso, del BOE. La sua ampiezza dň indicazione della non-specificitị $\mathcal{P}$ e si riduce a 0 per le probabilita̲yesiane. 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 $\geq 1/K$.

In calce: 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 $(0 \leq r \leq K)$ 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, $|\mathcal{F}| \leq \binom{K}{\lfloor \frac{K}{2} \rfloor}$.
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:

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

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

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

ossia

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

Ma allora:

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

□

Appendice: Probabilità bernoulliane

In questa sezione digressiva faremo un primo affondo, molto blando e formale, nella probabilità.

> Probabilità, casualità, determinismo sono parole da far tremare i polsi; per il loro significato che è largamente extra-matematico, vanno ben oltre il facile e formale discorso di "pesi" e "misure" che segue. Ai probabilisti impegnati sarà chiaro che lo schema proposto è (per ora!) bernoulliano, per cui le componenti dell'n-upla x possono esser viste come le prime n uscite di un processo stocastico stazionario e senza memoria (in termini familiari allo scienziato naturale: come i risultati osservati in n esperimenti replicati "nelle stesse condizioni"); i meno impegnati possono dare un'occhiata all'appendice. Jacob Bernoulli, autore dell'*Ars conjectandi* pubblicata postuma a Basilea nel 1713, è uno dei padri fondatori del calcolo delle probabilità.

`Le tre probabilità.`

> La parola probabilità si incontra in contesti che appaiono poco conciliabili.

`Probabilità combinatoria, o classica.`
La probabilità che esca un numero pari sul dado che sto per lanciare è di tre sesti, cioè di un mezzo, perchè gli eventi possibili sono sei, per ragioni di simmetria

sono equiprobabili, e quelli favorevoli sono tre. Il soggetto (lo sperimentatore, l'esperto) deve assumersi la responsabilità di saper distinguere i dadi costruiti "a regola d'arte" da quelli truccati.

`Probabilità empirica, o oggettiva, frequentista, statistica.`
La probabilità che il pezzo prodotto da questa macchina sia difettoso è del 3%, come ha mostrato la sperimentazione ripetuta e registrata un altissimo numero di volte. Si noti che non siamo di fronte al "valore esatto" della probabilità, ma solo a una sua *stima statistica*, sperabilmente accurata; il valore esatto è asintotico, raggiungibile solo dopo aver effettuato ... infiniti esperimenti. Di solito (con una certa *nonchalance* filosofica), la concezione empirica "divora" quella combinatoria: si suppone che nel caso di dadi costruiti "a regola d'arte", su cui si sono fatti esperimenti per centinaia d'anni in tutti gli angoli del globo, i valori limite possano considerarsi noti. Questa popolarissima "scappatoia filosofica" circola col nome pomposo di *legge empirica del caso.*

`Probabilità epistemica, o soggettiva.`
La probabilità che Israeliani e Palestinesi firmino l'accordo entro Pasqua è del 60%. In quest'ultimo caso non ci sono simmetrie né c'è la possibiliṭi ripetere gli esperimenti, non esistendo nessuna fantascientifica macchina del tempo che ci faccia ritornare centinaia di volte a Natale, diciamo, per ricontrollare ogni volta come vada poi a finire: qui c'è solo l'opinione di un esperto di politica disposto a scommettere sulla pace 6 contro 4 - l'obiezione casomai è come sia giunto a una valutazione numerica cosıprecisa. La probabilità epistemica ha la pretesa di essere "onnivora": in effetti se lo stato di conoscenza del soggetto è determinato "essenzialmente" dalla sua esperienza statistica (dagli esperimenti che ha svolto) la valutazione numerica "soggettiva" di una probabilità finisce col coincidere con quella "oggettiva" dell'empirista (nell'esempio precedente il soggettivista potrebbe ben puntare 3 contro 97 sull'uscita di uno scarto).

Dal punto di vista della "matematica astratta" la lacerazione filosofica è irrilevante, poichè, come si puømostrare, il calcolo delle probabilità combinatorie, oggettive e soggettive rimane lo stesso - i teoremi non cambiano.

> Il calcolo delle probabilità, che è lo strumento pĭtradizionale di gestione dell'incertezza (ma non l'unico, come vedremo), è nato ufficialmente nel 1654 con un memorabile scambio di sette lettere fra Blaise Pascal e Pierre de Fermat; in esse è risolto un problema di gioco d'azzardo futile solo per chi non sappia vedere la fertile metafora in esso celata. E' fin dai primordi che la probabilità si presenta scissa fra due concezioni che solo in un primo tempo i pionieri del nuovo calcolo riuscirono eclettticamente e fruttuosamente a contemperare: da una parte la probabilità *oggettiva*, legata a proprietà fisiche (a "dati di fatto"), come la simmetria di un dado o la regolarità statistica registrata in una serie di esperimenti, dall'altra la probabilità *soggettiva*, che misura il "grado di fiducia" (soggettivo ma razionale)

> del soggetto giudicante circa la possibile realizzazione di un evento, grado di fiducia che dipende dallo stato di conoscenze nel quale il soggetto giudicante si trova in quel momento.
>
> Entrambe le concezioni, quella oggettiva e quella soggettiva, rivendicano la stessa paternità: l'*Ars Conjectandi* di Jacob Bernoulli, che comparve postuma a Basilea nel 1713. Posta a buon uso nell'Ars Conjectandi, l'ambiguità col tempo divenne pesante, tant'è che agli inizi dell'Ottocento, ma senza gran successo, Siméon-Denis Poisson, anche lui illustre fisico oltre che illustre matematico, propose di usare due parole distinte, *probabilité* e *chance*, e di riservare la prima alla concezione epistemica, soggettiva, proprio come vuole la filologia. Una divagazione filologica non è in effetti fuori luogo: è interessante capire il perchè di un'ambiguità che all'inizio suggeriilluminanti analogie, ma che poi portò a fraintendimenti e ad abusi.
>
> Grosso modo fino al Seicento, il termine "probabile", affine in maniera ovvia a parole come "provabile" o "approvabile", era univoco e significava "approvato dall'autorità religiosa", "attestato dalle sacre Scritture". Non sorprende dunque che, senza contraddizione, i miracoli venissero considerati eventi probabili e insieme impossibili (*sic*). Come mai alla fine del Seicento il significato originale venne rapidamente obliterato e sostituito da due significati distinti e conflittuali? Secondo quanto affermano gli storici, ciò sarebbe dovuto a una ragione profonda, legata al ribaltamento delle fonti di autorità tradizionali, quelle religiose, che vennero soppiantate dall'autorità della "natura", natura che l'uomo interroga tramite il metodo sperimentale, e dall'autorità della "mente illuminata". Al tramonto dell'età dei lumi l'accordo fra natura (probabilità oggettiva) e mente illuminata (probabilità epistemica) si incrinò definitivamente. Tutto ciò può lasciar perplessi, tanto piěhe la probabilità, e con lei la statistica che le è ancella, hanno registrato e continuano a registrare successi pratici e applicativi che sono sotto gli occhi di tutti.

La *potenza cartesiana* n-esima $\mathcal{A}^n$ dell'alfabeto $\mathcal{A}$ è costituita dalle K^n *sequenze* di *lunghezza* n, o n-uple, costruite su $\mathcal{A}$. Alle n-uple e agli insiemi di n-uple in $\mathcal{A}^n$ associeremo ora dei "pesi", in un senso un po' vago ma suggestivo di questo termine. Partiremo dai "pesi atomici" delle lettere di $\mathcal{A}$, e poi specificheremo le regole per calcolare il peso di un'n-upla a partire dai pesi delle singole lettere che la costituiscono (*piano diacronico*), e il peso di un insieme di n-uple a partire dai pesi delle singole n-uple che lo costituiscono (*piano sincronico*).

> Usiamo il termine diacronico perchè troviamo suggestivo pensare che ciascuna n-upla sia generata via via da una *sorgente di informazione* dalla prima all'n-esima lettera componente; va da sé che, eccedendo in termini, stiamo preparando il terreno agli usi futuri di questi concetti. Il termine sincronico è usato piěhe altro per contrapposizione,

> come se tutte le n-uple dell'insieme A fossero state generate "in parallelo" da tante sorgenti quante sono le n-uple.

Piano istantaneo: I pesi P (le probabilità) delle K lettere sono K numeri non negativi che sommano ad 1: $p_i = P(a_i) = \text{Prob}(a_i) \geq 0$, $\sum_i p_i = 1$

> Ribadiamo che sommatorie "tacite" si intendono estese a tutti i valori dell'indice.

Piano diacronico: Se dagli n "atomi" si passa all'n-upla che ne è costituita, la probabilità bernoulliana ha comportamento *moltiplicativo*, nel senso che per ogni n-upla $\underline{x} = x_1 x_2 \ldots x_n$ si ha:

$$P^n(\underline{x}) = \text{Prob}(\underline{x}) = \prod_{1 \leq i \leq n} \text{Prob}(x_i)$$

Piano sincronico: Se si passa a un insieme di n-uple la probabilità bernoulliana ha comportamento additivo, nel senso che il "peso complessivo" di un insieme $A \subseteq \mathcal{A}$ di oggetti, di n-uple, è la somma dei "pesi" degli oggetti che lo compongono:

$$P^n(A) = \text{Prob}(A) = \sum_{\underline{x} \in A} \text{Prob}(\underline{x})$$

Esercizio. $H(P^n) = nH(P)$. (Per induzione su n.) La formula ben sottolinea il "carattere logaritmico" dell'entropia. Sul piano euristico: l'incertezza circa il risultato globale di n esperimenti "ripetuti nelle stesse condizioni" è la somma delle incertezze nei singoli esperimenti.

Complementi. L'additività sincronica implica le seguenti relazioni (con $m = P^n$) che faranno la gioia di chi ama la teoria della misura, visto che al loro interno riconoscerà gli assiomi delle misure su spazi finiti (sotto non abbiamo specificato n, poichè qui la dimensione è irrilevante; $\mathcal{A}lU$ è l'insieme pieno; nel caso di P^n, $\mathcal{A}lU = \mathcal{A}^n$): $m(A) \geq 0\ ,\quad m(\emptyset) = 0$

$$m(A \cup B) = m(A) + m(B) - m(A\mathcal{A}pB)$$
$$m(A \cup B) \leq m(A) + m(B)$$
$$A\mathcal{A}pB = \emptyset \Rightarrow m(A \cup B) = m(A) + m(B)$$
$$m(\overline{A}) = m(\mathcal{A}lU) - m(A)$$

Una formula speciale (e molto amichevole) per le probabilità bernoulliane (che sono moltiplicative sul piano diacronico e additive su quello sincronico) la si ha nel caso dei prodotti cartesiani(□ !):

$$P^{nm}(A \times B) = P^n(A)P^m(B)$$

Cenno: dimostra prima la $P^{nm}(\underline{x} \times B) = P^n(\underline{x})P^m(B)$, poi somma su $x \in \mathcal{A}$.

> L'amichevole formula riflette la mancanza di memoria dei processi bernoulliani (indipendenza delle n componenti aleatorie).

Come corollario si trova $P^n(\mathcal{A}^n) = 1$ (per induzione). Dunque la probabilità è, come si usa dire, una misura *normata a* 1.