LINEAMENTI DI CALCOLO DELLE PROBABILITA', OVVERO, OTTIMIZZAZIONE DELLE DECISIONI IN CONDIZIONI D'INCERTEZZA

1. Introduzione

Scrivere queste pagine è un compito al quale ci accingiamo malvolentieri, una volta ancora da "grafoman[i] controvoglia":

"Sostengo da sempre che (salvo casi di patente masochismo), non esiste quella 'gioia di scrivere' di cui parlano quasi solo coloro o che non praticano questo esercizio o che non trovano poi chi stampi le loro scritture. Esiste, semmai, la 'gioia di aver scritto': il sollievo di avere finito, di essersi liberati da un'attività che non e' affatto 'naturale', che si rivela spesso un peso, un tormento. E non mi si accusi di incoerenza: perché, allora, avrei passato la vita a scrivere, invece di praticare mestieri meno penosi? Ho la risposta pronta: tanto sono restio alla scrittura, altrettanto sono goloso di lettura. Dunque, per ciò che davvero mi interessa, non avendo trovato i libri che avrei voluto leggere, ho ceduto a una *raptus* di presunzione, tentando di farmeli da solo [...] questo che leggo non mi appaga; conosco i miei gusti, so quel che cerco; dunque, vedo di prepararmi io stesso quel che mi piacerebbe leggere..." (Vittorio Messori, dalla Prefazione a *Il quadrato magico*, di Rino Cammilleri, Ed. Rizzoli, Milano, 1999).

In effetti le presenti sommarie riflessioni sui fondamenti del Calcolo delle Probabilità hanno origine da **due** irritazioni.
La **prima**, una delle solite battute che vengono attribuite al "grande" Bertrand Russell: "*Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means*" (da una lezione da lui tenuta alla Beacon Hill School nel 1929, la scuola era stata fondata dallo stesso Bertrand Russell e dalla moglie Dora nel 1927), considerazione che fa il paio con la sua famosa definizione della matematica: "*mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true*" (in "Mathematics and the Metaphysicians", contenuto in *Mysticism and Logic and Other Essays*, 1910, 1917).
Vedremo nel seguito come la teoria della probabilità non sia affatto una materia poco chiara, anzi, intesa come una **logica dell'incertezza**, sia sostanzialmente una soltanto, niente "scuole" o "impostazioni". Si parla infatti di approccio classico, frequentista, soggettivista, assiomatico, etc., tutte distinzioni che vengono **dopo**, ossia quando si tratterà di assegnare un valore di probabilità alle diverse modalità con cui può verificarsi un evento. Volendo anticipare un poco il nostro punto di vista, accettando quindi un modo di procedere che vuole andare subito al "dopo", diciamo che ci muoveremo nell'ambito di un'ottica bayesiana (da Wikipedia: "anziché la frequenza o la propensione di qualche fenomeno, la probabilità viene interpretata come aspettazione razionale rappresentante uno stato di conoscenza o come quantificazione di una

convinzione personale") e quindi soggettivista: l'attribuzione di una probabilità da parte di un dato osservatore ω ad una modalità di un certo evento non può essere altro che il frutto della valutazione dello stato della conoscenza di ω rispetto a quell'evento, e se si vuole dare un po' di senso alla battuta di Russell bisogna riconoscere che non si tratta altro che di un tentativo di nobilitare la nostra permanente incertezza attraverso una dignitosa veste matematica (non ci si può esimere dal rilevare come la situazione, almeno a volte, faccia andare la mente al *latinorum* di Don Abbondio). Una quantificazione spesso non ha infatti altro scopo che quello di introdurre una relazione di **preordine** tra le modalità di un evento, con i numeri è più bello, e appare più "scientifico".

[Un **preordine** è una relazione binaria interna definita su un dato insieme X che sia riflessiva e transitiva, ma non necessariamente antisimmetrica. In un preordine, $x < y$ per due elementi x, y di X significherà $x \leq y$ e x non $\geq y$. Il preordine è alla base di diverse costruzioni della matematica, preordine tra segmenti di una retta (teoria della misura, o dei numeri prima razionali e poi reali), preordine tra insiemi finiti (aritmetica), etc., incluso un preordine in calcolo delle probabilità che si limita a decidere tra due modalità di uno stesso evento quale delle due è maggiormente probabile dell'altra, non escludendosi appunto il caso che esse siano per noi ugualmente probabili.]

La **seconda**, un accenno di Federico Faggin nel corso di una conferenza alla definizione di quantità d'informazione secondo Shannon: "Cosa vuol dire informazione astratta che non vuol dire niente? [...] Shannon ha definito l'informazione come essenzialmente è il cologaritmo della probabilità di un evento, cologaritmo della probabilità di un evento si chiama informazione. In realtà la chiama quantità di informazione, comunque questa è la definizione, e questa probabilità non ha nulla a che fare con l'esperienza, con la conoscenza di quell'informazione, è semplicemente sufficiente riconoscere l'evento" (F. Faggin, L. Rumor: IRRIDUCIBILE La coscienza, la vita, i computer e la nostra natura, dal minuto 28':
https://www.youtube.com/watch?v=C5fgYfNQJp0&t=78s)

Lascio ai lettori (presumibilmente assai pochi - ancora una volta, nel corso della vita, mi trovo a dover dire come Carmen: "*Je chante pour moi même*") la responsabilità di decidere se la precedente incomprensibile spiegazione (l'informazione è la stessa cosa della sua quantità? un evento, o la sua probabilità, costituiscono informazione, o è piuttosto l'informazione che viceversa determina la probabilità di un evento? ne riparleremo) rispetta il proposito espresso da Galileo quando cercò di illustrare il significato di proporzione geometrica, ossia, come diremmo oggi, il concetto di numero reale: "avendo il lettore concepito **già** nell'intelletto che cosa sia la proporzione fra due grandezze [...] **mi sforzerò di secondare con la difinizione delle proporzioni il concetto universale degli uomini anche ineruditi nella geometria**" (Principio

di giornata aggiunta - Giornata quinta, "Sopra le definizioni delle proporzioni d'Euclide", contenuto nei *Discorsi e Dimostrazioni matematiche intorno a due nuove Scienze*, Leiden, 1638).
Come dire che procederemo qui secondo l'intento espresso nel titolo dell'opera fondamentale di Georges Boole, ***An Investigation of the Laws of Thought*** *on Which are Founded the Mathematical Theories of Logic and Probabilities* (1854), e sempre rispettando l'ammonimento di Cartesio (1637), secondo il quale ogni *Discours* deve avere come scopo principale quello di *bien conduire sa raison*...
Il che val pure a significare che quella che seguirà è la descrizione dell'attività del **mio proprio** intelletto quando affronta le questioni di cui in titolo, ed è sempre opportuno in siffatti contesti precisare, per evitare scontate critiche di relativismo, o di soggettivismo esasperato fino al solipsismo, che allo schema filosofico dianzi sommariamente delineato va aggiunto un *principio di universalità*: ossia, descrivo il mio proprio intelletto perché solo questo posso fare, ma sono invero persuaso che così come funziona il mio intelletto funzionano anche tutti gli altri intelletti, *ergo*, sto descrivendo non solo **un** intelletto ma **tutti** gli intelletti, cioè **l'intelletto** umano con l'articolo determinativo.

Ciò doverosamente premesso, la prima questione che dobbiamo affrontare è quella che concerne la definizione di **evento** (e poi di **situazione**), il che ci conduce a qualche commento in ordine al concetto fondamentale di "definizione" in matematica. E' infatti evidente come spesso ci si trovi di fronte a semplici giri di parole, sinonimi, cioè a procedimenti sostanzialmente tautologici, dimenticando che insieme alle possibilità di una **definizione intensiva** (quindi con tutti i limiti di qualsiasi **linguaggio**), esistono anche le possibilità di **definizioni ostensive**, che si riferiscono quindi direttamente al pensiero (ossia, senza l'intermediazione del linguaggio), tutto ciò tenendo ben presente la tricotomia fondamentale dell'esperienza umana:
reale - pensato - parlato.
Alla complessa questione abbiamo posto spesso attenzione in altri analoghi contesti, tenendo ben presente una famosa osservazione di Kant, "La colomba leggera, mentre nel libero volo fende l'aria di cui sente la resistenza, potrebbe immaginare che le riuscirebbe assai meglio volare nello spazio vuoto di aria" (*Critica della ragione pura*, Introduzione), e ricordando come esempio paradigmatico che è in effetti fondato ritenere che le famose definizioni euclidee vengano utilizzate come un mezzo per indicare (alludere a) un oggetto esistente, ancorché ideale (Max Simon le paragona alle indicazioni sulla nomenclatura che il maestro di bottega fornisce all'apprendista, quando gli dice mostrandogli gli attrezzi: "Questa è la pialla, questa è la sega" - citazione da Attilio Frajese, *La matematica nel mondo antico*, Ed. Studium, Roma, 1951, p. 79).

Insomma, tenuto conto di tutto ciò che precede, cosa dovremo intendere con il termine "evento"? Si tratta di un evento, appunto, o avvenimento, o fatto, etc., che si può riferire al futuro, ma anche al passato*, ed essere attinente al reale ma anche solo al pensato. E' inevitabile che un evento venga descritto attraverso un linguaggio (anche comunicando tra sé e sé), ma dobbiamo accettare che un **evento** sia una classe di equivalenza di sue descrizioni linguistiche, modulo un non precisamente formalizzabile concetto di *isomorfismo semantico* (il che rappresenta un'ulteriore facile confutazione della proposta identificazione uomo-macchina, che da Turing in avanti è alla base del progetto della cosiddetta Intelligenza Artificiale). Ci si basa cioè sulla capacità dell'intelletto di saper decidere se due descrizioni sono **la stessa cosa**, oppure no.
Ciò detto, una tale definizione di evento non ci sembra sufficiente per poter procedere verso la nostra mèta, in quando ad un evento deve essere anche associata una specifica **analisi** di esso, ossia una dettagliata descrizione delle possibili **modalità** che dell'evento si ritengono per noi interessanti. Anche qui si potrà introdurre un'opportuna nozione di isomorfismo, sicché arriviamo finalmente a quello che riteniamo l'oggetto fondamentale del nostro studio, ossia alla definizione di **situazione**. Questa sarà costituita da una coppia ordinata il cui primo elemento è un evento, ossia una classe di equivalenza di descrizioni semanticamente isomorfe dell'evento, mentre il secondo sarà una classe di equivalenza di analisi dell'evento.

Note al Cap. 1

* Ci piace nell'occasione citare la seguente frase:
"Per quanto riguarda la probabilità che questi eventi abbiano mai avuto luogo"
https://it.wikipedia.org/wiki/Vitello_d%27oro,
a dimostrazione che di probabilità - ovvero di incertezza - è legittimo parlare anche a proposito di eventi storici sulle cui modalità di effettiva realizzazione abbiamo incertezza. L'autore si è occupato a lungo del caso della scomparsa di Ettore Majorana, e quindi dell'assegnazione di probabilità - ovviamente soggettive, in tale soggettività rientrando anche la facoltà di rifiutare quanto altri ritengono invece certo** - alle diverse possibilità che si aprono alla mente nella discussione del caso. Proporremo nel seguito un "grafo" universale relativo al problema di qualsiasi scomparsa.

** E qui ci pare opportuno riportare una nostra non breve analisi del concetto di **verità**, ossia dell'opposto almeno apparente di **incertezza**, redatta in occasione di una discussione sul caso di S. Giuseppe da Copertino.
http://www.cartesio-episteme.net/ep8/ep8-sebast.htm
Le verità si possono ripartire in **tre** classi fondamentali.
1 - Le verità di natura che potremmo dire **logica**, che in senso kantiano comprendono i *giudizi sintetici a priori*, su cui si edifica per esempio la matematica, e i *giudizi analitici* che regolano i processi di deduzione logica, tanto in matematica come altrove.

2 - Le verità di natura **sperimentale**, ovvero, ancora in senso kantiano, i *giudizi sintetici a posteriori*, che vengono desunti dall'osservazione della "realtà materiale" (per Kant la "realtà materiale" costituisce il *fenomeno*, dal momento che in ogni caso essa è percepibile soltanto attraverso la mediazione delle forme pure dell'intelletto, spazio e tempo). Esse consistono sostanzialmente nella descrizione di processi naturali che si ripetono identici nelle medesime condizioni, e nella formulazione di "leggi" generali che sono determinate attraverso un procedimento d'*induzione*, dei cui limiti epistemologici bisogna sempre essere ben consapevoli.
3 - Le verità di natura **storica**, che attengono anch'esse al campo dei giudizi a posteriori, con la differenza che si tratta di fenomeni unici, non (almeno esattamente) ripetibili, oggetto esclusivamente di racconti, testimonianze, destinati alla "memoria".
E' chiaro che nella presente circostanza dobbiamo esaminare eventuali verità della terza categoria, che offrono gravi difficoltà di accertamento all'investigatore, e impervi ostacoli di trasmissione interpersonale (intervenendo qui, inoltre, gli ordinari inevitabili fraintendimenti linguistici: "*le language est source de malentendus*", come rammenta Saint-Exupéry in *Le Petit Prince*), ma non per questo possono essere espunte dall'ambito di ciò a cui è lecito attribuire le specificazioni "vero" o "falso". Se è arduo tracciare una netta linea di demarcazione tra "fatti" e "discorsi sui fatti", i primi restano in ogni caso gli elementi essenziali della trama costitutiva della storia, e con essi bisogna fare i conti - obbligo che vale pure per i mistificatori, gli esperti in disinformazione. Tutti i "creatori di storie" agiscono di solito innestando elementi fantastici su quelli reali, e favorendo resoconti ad arte parziali, che consentano le desiderate interpretazioni di comodo. Rimangono però, intorno a un evento realmente verificatosi, e significativo al punto da provocare conseguenze, una serie di concomitanti evidenze, indizi (per ulteriori riflessioni sulla necessità del metodo indiziario nella ricerca storiografica, si veda: http://www.cartesio-episteme.net//st/indiz.htm), ed è lecito allora anche "accontentarsi" di poter solo **intravedere la verità**. Bernard Fay (nella Prefazione a *La Franc-Maçonnerie et la révolution intellectuelle du XVIIIe siècle*, Ed. de Cluny, Parigi, 1935), scrive che è possibile soddisfare "la passion de comprendre" dal momento che "les hommes n'ont pas détruit tout ce qu'ils croyaient détruire ni caché tout ce qu'ils voulaient dissimuler; ce désordre permet a l'historien d'entrevoir parfois la vérité". Così, in questo senso, quando sia naturale presumere un deliberato intervento correttore dell'uomo sulla semplice verità fattuale, il lavoro dello storico può assomigliare a quello del decifratore: il suo compito è di separare il grano dal loglio, di scavare nelle zone d'ombra, sospette al pari delle zone di luce eccessiva, pervenendo semmai a individuare un ventaglio di ragionevoli alternative, e a descriverne la maggiore o minore probabilità. Un apprezzabilissimo esempio dell'applicazione di tale metodo ci sembra offerto da Flavio Barbiero, nel suo ottimo *La Bibbia senza segreti*, per il quale rimandiamo senz'altro alla presentazione che ne fu fatta in Episteme N. 2
(http://www.cartesio-episteme.net/episteme/epi2/ep2barb2.htm).
Di fronte alle verità del terzo tipo si riscontra un ampio spettro di atteggiamenti. E' frequente oggi imbattersi, specialmente in ambito accademico, in "scettici" *part time*, che si mostrano dubbiosi dello stesso concetto di verità storica (evidenziando così

l'ovvio lato soggettivo di ogni tentativo di storiografia, come messo bene in evidenza da Hegel: "E' giusto esigere che la storia, quale ne sia l'argomento, racconti i fatti senza parzialità, senza pretendere di avvalorare interessi o scopi particolari. Ma tale esigenza è un luogo comune che approda a ben poco, giacché la storia di un argomento è necessariamente collegata in modo strettissimo all'idea che ci facciamo di esso. Questa fissa già in precedenza che cosa si considera importante e conveniente per l'argomento prescelto, e siffatto rapporto tra quanto è accaduto e lo scopo che ci proponiamo porta seco la selezione dei particolari da raccontare, il modo d'interpretarli, i punti di vista sotto i quali collegarli", *Lezioni sulla Storia della Filosofia*), ma non di quelle (e quindi dei connessi "giudizi di valore") che costituiscono il fondamento incriticabile (se non a prezzo di sgradevoli contestazioni) della società a cui detti "scettici" appartengono. Ci piace concludere il presente intermezzo accennando al parere più accettabile, e "umano", espresso da Marguerite Yourcenar ("Taccuini di appunti" per le *Memorie di Adriano*, 1971): "Tutto ci sfugge. Tutti. Anche noi stessi. La vita di mio padre la conosco meno di quella di Adriano. La mia stessa esistenza, se dovessi raccontarla per iscritto, la ricostruirei dall'esterno, a fatica, come se fosse quella d'un altro. [...] Il che non significa affatto, come si dice troppo spesso, che la verità storica sia sempre e totalmente inafferrabile; accade della verità storica né più né meno come di tutte le altre: ci si sbaglia, *più o meno*".]

2. Il grafo associato ad una situazione e le prime probabilità a priori, e a posteriori, relative e assolute

$$\underset{\text{osservatore}}{\omega} \longrightarrow \underset{\text{situazione}}{\boxed{\mathcal{S}}} \longrightarrow \text{decisione}$$

Il precedente disegnino rappresenta l'oggetto della nostra indagine. È da subito chiaro che bisognerà prendere in considerazione l'analisi $\mathcal{A}$ che fa parte di $\mathcal{S} = (\underset{\text{evento}}{\mathcal{E}}, \underset{\text{analisi}}{\mathcal{A}})$.

$\mathcal{A}$ individua le possibili modalità con cui può avverarsi l'evento $\mathcal{E}$, diciamole

$\mu_1 \; \mu_2 \; \dots \; \mu_n$ (ove $n \geq 2$ per ipotesi).

Si illustra in tale modo un primo livello di analisi, al quale può seguire un secondo livello:

$$\begin{array}{ccc} \mu_1 & \mu_2 & \dots \\ \mu_{(1,1)}\,\mu_{(1,2)} \cdots \mu_{(1,n_1)} & \mu_{(2,1)} \cdots \mu_{(2,n_2)} & \end{array}$$

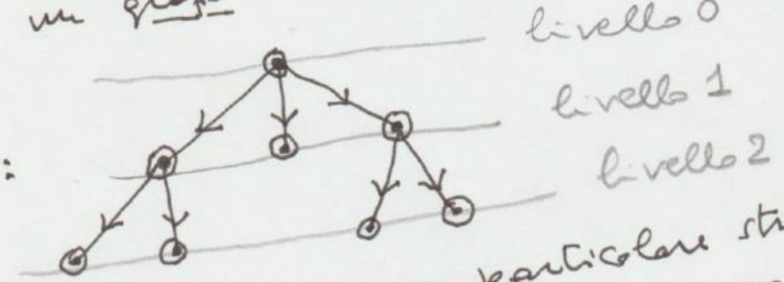
modalità mutuamente escludentisi

etc.

Si noti che per esempio n_1 può essere uguale a 0, vale a dire quando non siano indicate sotto-modalità per la modalità μ_1, e che comunque se $n_1 \geq 1$ allora deve anche essere $n_1 \geq 2$ per ipotesi.

In definitiva, possiamo descrivere la situazione con un grafo del seguente tipo:

ordine parziale = relazione binaria interna che sia riflessiva, antisimmetrica e transitiva.

Γ : 

che corrisponde ad una particolare struttura di ordine parziale sull'insieme di modalità e sottomodalità di $\mathcal{E}$, più un elemento minimo assoluto che si dice radice del grafo (al quale si pensa come un albero rovesciato; l'ordine si legge dall'alto verso il basso, le relazioni $x \leq x$ riflessive non vengono indicate).

EVITEREMO SEMPRE NEL SEGUITO DI RIPORTARE ESPLICITAMENTE LE FRECCINE $\downarrow$.

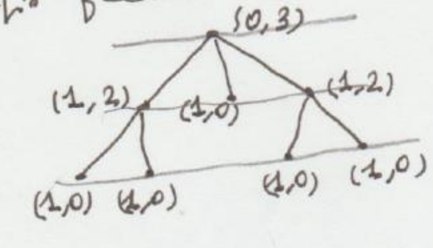
Il grafo precedente ha altezza $h = 2$, un unico elemento minimo assoluto, 5 elementi massimi relativi (8 nodi (1 al livello 0, 3 al livello 1, 4 al livello 2)). Il grafo ha la proprietà che per $\forall\, x, y \in \Gamma$, $x < y \Rightarrow \exists$ un unico cammino ordinato da x ad y (il che implica che Γ non ha cicli).

Ad ogni nodo si può associare una coppia ordinata di numeri naturali $\neq 0$, che si dice la valenza del nodo (o modalità):

$v(\nu) = (a, b)$ — a = n° dei lati in entrata da ν; b = n° dei lati in uscita da ν

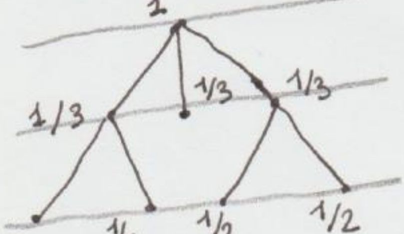
Nell'esempio precedente:

Γ :



Si noti che, ad eccezione della radice, tutti i nodi hanno 1 come primo elemento della relativa valenza. I massimi relativi hanno 0 come secondo elemento della valenza

Passiamo subito dal disegno di cui sopra al seguente:



(definizione pascaliana di probabilità)

In tale disegno abbiamo già voluto introdurre un particolare tipo di probabilità, che diremo probabilità astratta o aritmetica relativa, diciamola p_α. p_α sarà una funzione da Γ (dal sostegno di Γ) ad $I = [0,1] \subset \mathbb{R}$ (ci capiterà di usare quasi sempre nel seguito numeri razionali, ma introdurre qui $\mathbb{R}$ non stona particolarmente.

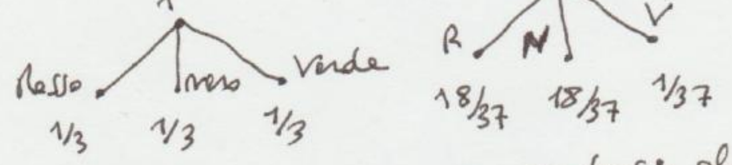
La definizione di p_α è ovviamente la seguente:

$p_\alpha(\mu) =$ $\rightarrow 1$ se μ è la radice del grafo

$\rightarrow 1/a$ se la valenza del nodo precedente di μ (che potremo indicare con simbolismo espressivo con $\mu - 1$) è $(1, a)$.

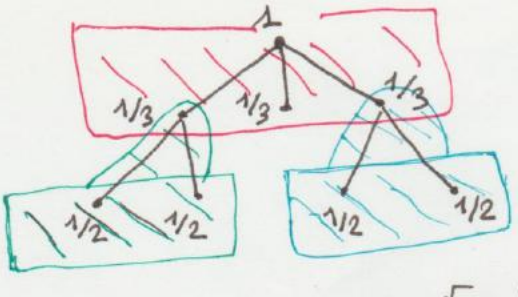
Naturalmente, potrà accadere che sin da subito p_α si riveli poco interessante per i nostri fini, in quanto tale probabilità astratta risulta ben diversa da quella effettiva, che diremo p_e.

Esempio, il colore che può uscire alla roulette. Abbiamo una tricotomia, in cui p_α vale $(1/3, 1/3, 1/3)$, mentre p_e vale $(18/37, 18/37, 1/37)$:



Un raffronto tra p_e e p_α ci conduce al concetto di non-equiprobabilità delle modalità della situazione in esame, un concetto sul quale torneremo.

Notiamo subito invece che $\forall$ nodo dell'albero che non sia uno dei massimi relativi è la radice di un sotto-albero che si può dire una sezione di Γ:



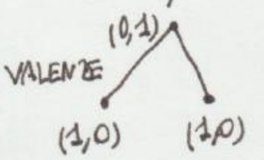
Ecco le 3 sezioni del nostro albero-esempio.

Si noti che ovviamente, $\sum_\mu p_\alpha(\mu) = \sum_\mu p_e(\mu) = 1$

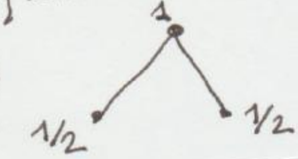
se la somma è estesa alle sole modalità che fanno parte della medesima sezione, e diverse ovviamente dalla radice della sezione (è questione di gusti decidere se la radice debba essere inclusa nella sezione oppure no).

Il grafo Γ insieme con la sua funzione di probabilità effettiva relativa $p_e : \Gamma \to \mathbb{R}$ descrive completamente la situazione che vogliamo analizzare. Due diverse situazioni $\mathcal{S}$ ed $\mathcal{S}'$ che producono a grafi associati $\Gamma(\mathcal{S})$ e $\Gamma(\mathcal{S}')$ che siano isomorfi e con corrispondenti funzioni di probabilità effettive relative potranno dirsi **equivalenti**.

Così, sono equivalenti (per esempio) le due situazioni facenti capo alla dicotomia standard



$\Rightarrow$



p_e

SITUAZIONI SIMILI!
NEL CASO DEI SOLI GRAFI ISOMORFI

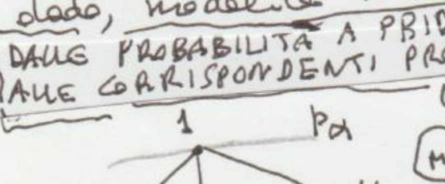
che si riferiscono al lancio di una moneta, modalità testa o croce, oppure al lancio di un dado, modalità numero uscito ≤ 3, numero ~~uscito~~ ≥ 4.

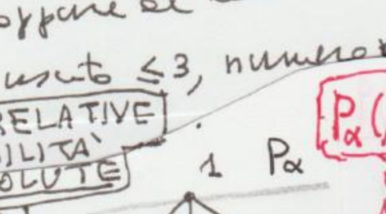
DALLE PROBABILITÀ A PRIORI RELATIVE ALLE CORRISPONDENTI PROBABILITÀ ASSOLUTE

P_α, P_e MAIUSCOLE

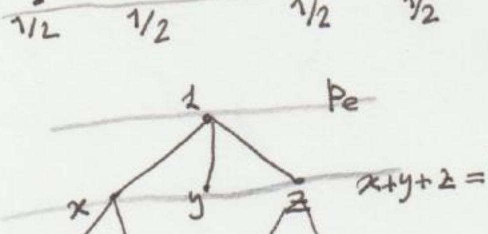
$$P_\alpha(\mu) = p_\alpha(\mu)\, P_\alpha(\mu-1)$$

indichiamo con $\mu-1$ l'antecedente di μ (se non è la radice)



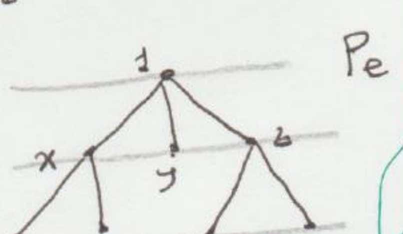


INSOMMA, $\exists$ 6 DIVERSI TIPI DI PROBABILITÀ, CHE DIVENTERANNO 12 NEL SUCCESSIVO CAPITOLO SULLE COMPATTIFICAZIONI



$x+y+z=1$

$t+u=1$ $\quad$ $v+w=1$



Ovviamente p_e, P_e, probabilità effettive a priori relative ed assolute, potranno essere sostituite da probabilità p, P, a posteriori, relative ed assolute, in corrispondenza ad una variazione dello stato dell'informazione $\mathcal{I}(\omega,\mathcal{S},t_0) \Rightarrow \mathcal{I}(\omega,\mathcal{S},t)$ dal tempo t_0 al tempo t. Ce ne occuperemo in esteso nel cap. 5.

[redacted]

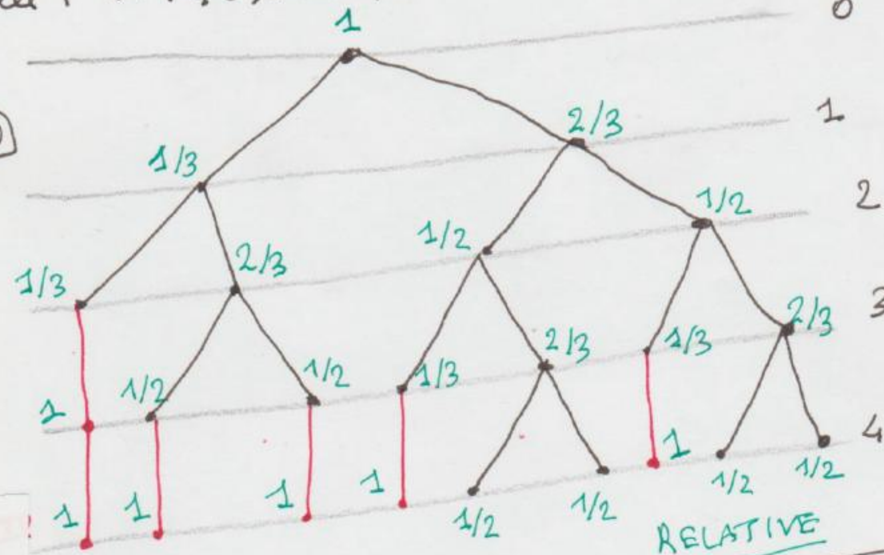
(11)

L'introduzione delle probabilità assolute merita forse qualche ulteriore spiegazione. Al 1° livello ovviamente probabilità relative ed assolute coincidono, si procede poi dall'alto verso il basso utilizzando la regola

$$P(\mu) = p(\mu)\, P(\mu-1) \qquad (\mu \neq \text{radice},\ \mu-1 = \text{antecedente di } \mu)$$

(regola che permette di ricostruire le p dalle P se P≠0, sulla questione dello 0 torneremo nel cap. 5)

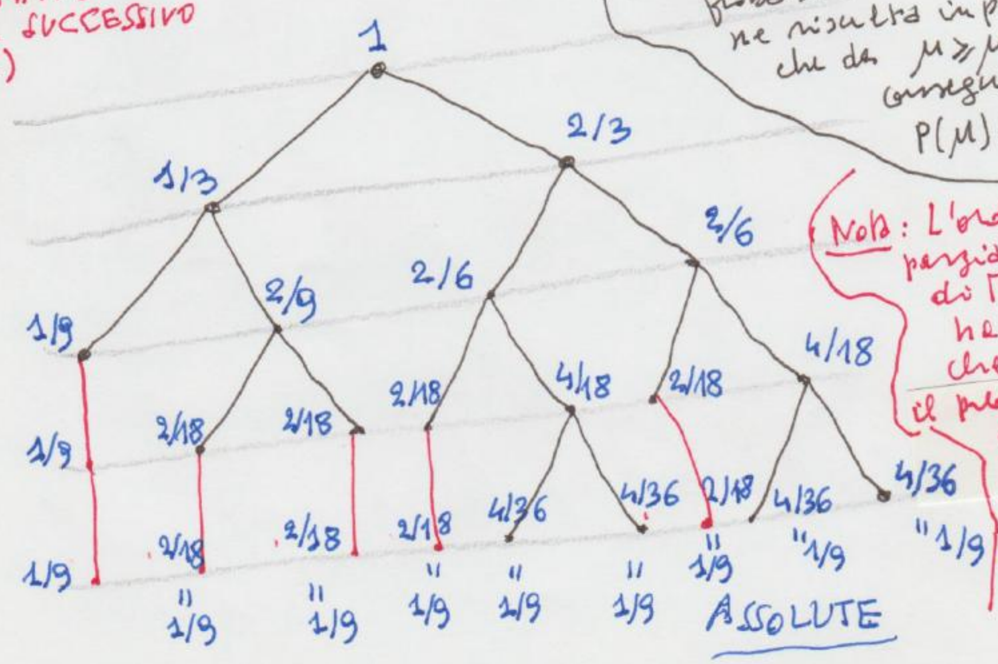
ESEMPIO



anche se $x \leq y$ in $\Omega \Rightarrow y \leq x$ in Π.

IN ROSSO LA COMPATTIFICAZIONE DEL GRAFO (VEDI IL SUCCESSIVO CAP. 3)

Nota: Poiché tutte le probabilità sono ≤ 1, ne risulta in particolare che da $\mu \geq \mu'$ consegue $P(\mu) \leq P(\mu')$



Nota: L'ordine parziale Ω di Γ NON ha a che fare con il preordine duale Π indotto su Γ dalle P.

le probabilità a posteriori, di

natura eminentemente soggettiva,

dipenderanno dallo stato della conoscenza,

o dell'informazione, di ω rispetto ad

S in un determinato tempo t, in simboli:

$$\mathcal{I}(\omega, S, t),$$

per cui le precedenti possono essere riferite

ad uno stato dell'informazione in un tempo

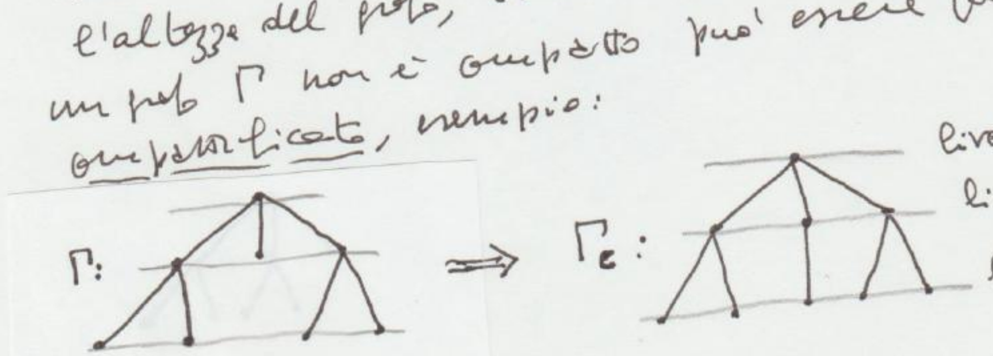
iniziale t_0, informazione astratta o effettiva,

riassumendo: $\mathcal{I}_a(\omega, S, t_0)$, $\mathcal{I}_e(\omega, S, t_0)$, $\mathcal{I}(\omega, S, t)$.

3. Grafi compatti e compattificazioni

Il grafo esempio da cui siamo volutamente partiti ha la sgradevole caratteristica che non tutti i suoi elementi massimali (relativi) hanno la stessa distanza dalla radice. Li hanno infatti distanza 2, 1 distanza 1.

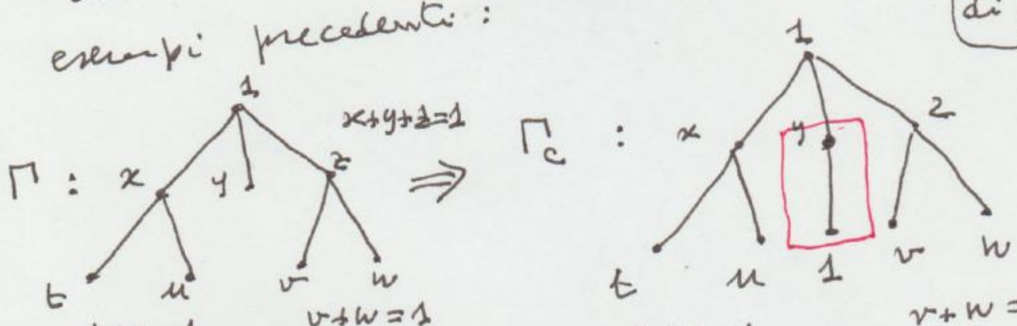
Un grafo in cui ciò non accade, ovvero in cui tutti gli elementi massimali hanno dalla radice la stessa distanza, che coincide allora con l'altezza del grafo, si dirà compatto. Se un grafo Γ non è compatto può essere però facilmente compattificato, esempio:



dove così facendo abbiamo rinunciato alla proprietà che ∀ nodo del grafo ~~che non~~ sia un massimo relativo ha valenza (1,b) con $b \geq 2$, nell'esempio precedente abbiamo infatti una valenza (1,1). +1

Ma tant'è, possiamo adesso estendere facilmente probabilità relative o assolute da Γ a Γ_c. Con riferimento agli esempi precedenti:

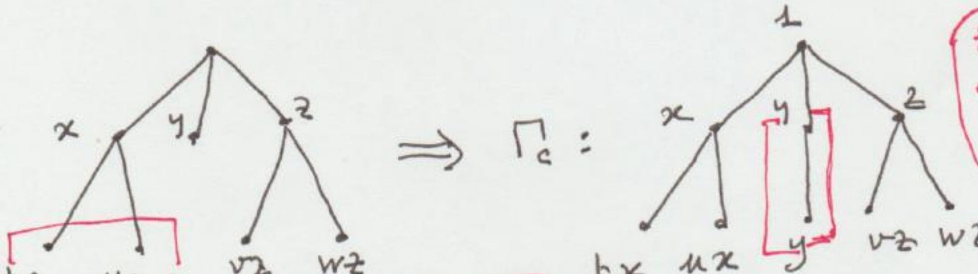
(manteniamo per esse lo stesso simbolo di partenza)



RELATIVA

(la Σ delle probabilità relative su ogni singola sezione del grafo, inclusa quelle con un solo ramo, è uguale ad 1)

ASSOLUTA



la somma di tutte le probabilità dello stesso livello è uguale ad 1.

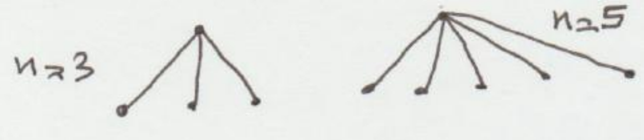
la somma $tx+ux$ è uguale alla probabilità della radice della sezione e adesso ovviamente

$tx + ux + y + vz + wz = 1$!

$x(t+u) + y + z(v+w)$

Abbiamo fin qui tralasciato di introdurre un'altra importante caratteristica dei nostri grafi, ovvero la loro complessità K = numero degli elementi massimi relativi, nel caso precedente K=5. Naturalmente, possiamo introdurre una nozione di complessità per ∀ livello di un grafo compatto.

Decomposizione di un grafo in compatto, h n-tomie, nel caso precedente:



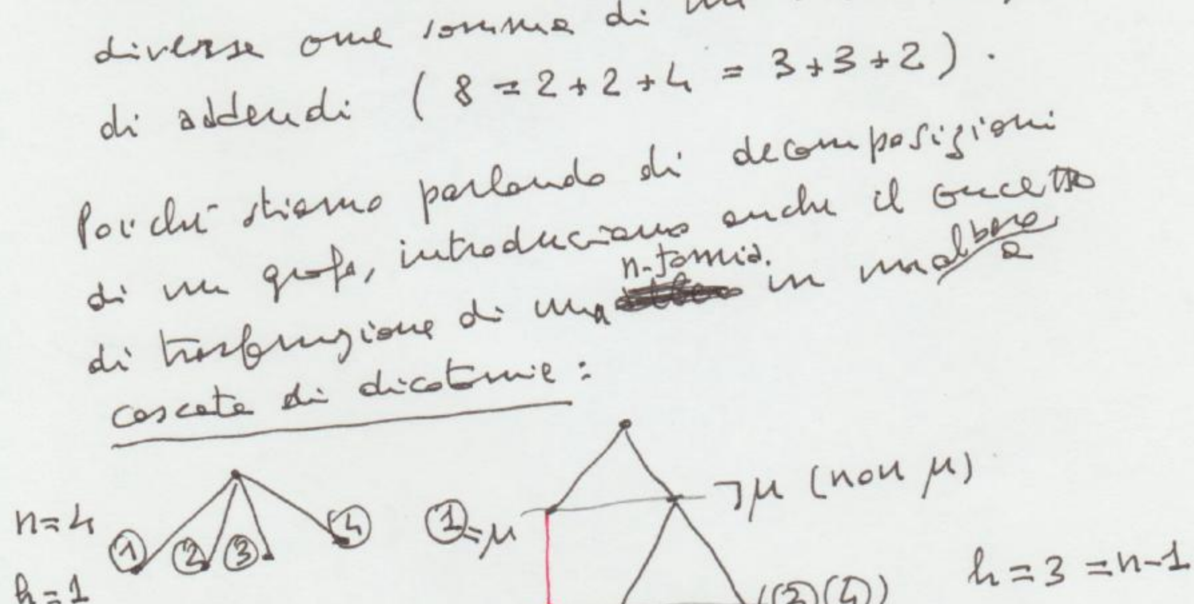
h = 2 altezza

k = 5 complessità (al secondo livello)

3 complessità al primo livello

Si noti esplicitamente che, decomposto il grafo, non è possibile univocamente ricomporlo, in quanto uno stesso numero naturale può scriversi in maniere sostanzialmente diverse come somma di un numero fissato di addendi (8 = 2+2+4 = 3+3+2).

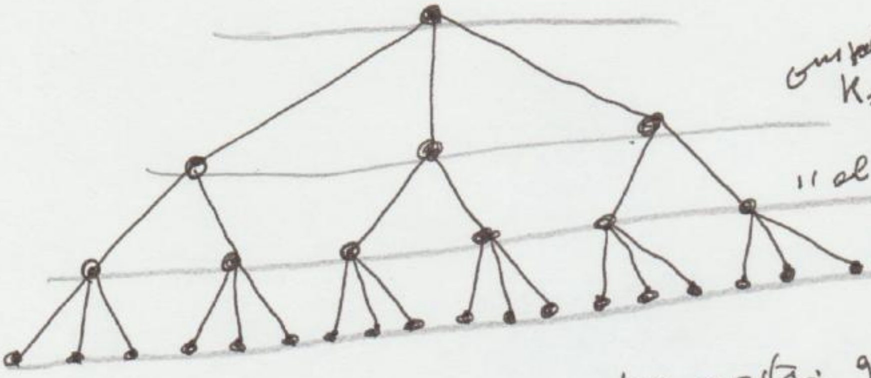
Poichè stiamo parlando di decomposizioni di un grafo, introduciamo anche il concetto di trasformazione di una n-tomia in un albero a cascata di dicotomie:



In rosso, il relativo grafo compatto.

Per chiudere questo capitolo, val la pena di introdurre la nozione di grafo completo di una certa altezza 2 e di complessità originarie ($M_1, M_2, \ldots, M_k$) -

Esempio, $h=3$, complessità originaria $(3,2,3)$ (14)+1



complessità al 1° livello $K_1 = M_1 = 3$

" al 2° $K_2 = M_1 M_2 = 6$

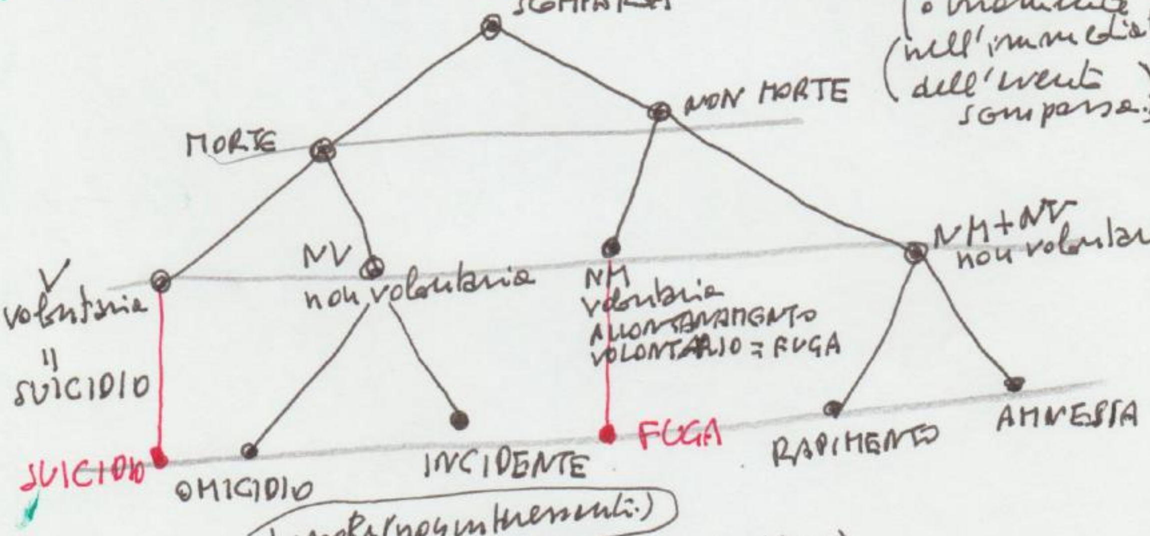
" al 3° $K_3 = M_1 M_2 M_3 = 18$

Manifestamente, ciascuno dei nostri grafi (alberi) è un sottografo di un opportuno grafo completo.*

Nota finale – Ecco come fornemo il grafo (non completo) che rappresenta l'analisi logica di qualsiasi tipo di scomparsa di una persona:

* Il che NON significa che le possibilità in che restrizioni di ritroveremo delle dall'uno all'altro grafo.

(Queste due modelli ovviamente nell'immediatezza dell'evento scomparsa.)

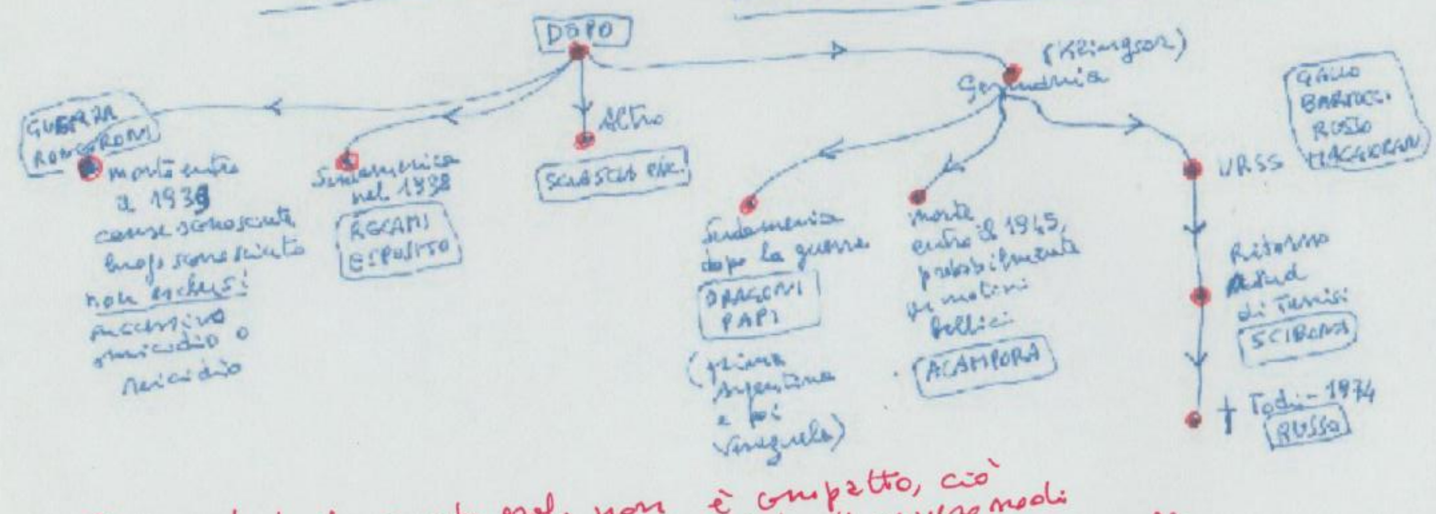


Oltre alle banali (non interessanti) $p_a = p_e$ e $P_\alpha = P_e$, spetterà all'investigatore/inquirente introdurre le probabilità a posteriori p, P che dipenderanno ovviamente dallo stato della conoscenza $\mathcal{I}(w, \mathcal{S}, t)$.

Altezza $h=3$ complessità $K_1=2$, $K_2=4$, $K_3=6$.

Esempio particolare, il grafo della scomparsa di Ettore Majorana, con cui si arricchisce il grafo "universale" precedente.

È ormai chiaro che la scomparsa di EM fu un allontanamento volontario, per ragioni del tutto personali, private. Cosa accadde di lui DOPO è oggetto di IPOTESI.
Esclusi omicidio o suicidio almeno nell'imminenza della scomparsa.



Nota: È manifesto che questo grafo non è compatto, cioè nonostante vi appaiono dei "rami solitari", ovvero nodi che hanno valenza in uscita = 1. Per renderlo più conforme alla nostra impostazione, basterà affiancare a ciascuno di questi rami solitari un secondo ramo parallelo riferito genericamente ad ALTRO.

È ovvio che in questo caso p_α, p_e, P_α, P_e, non servono a niente, bisogna introdurre p, P, e questa è una questione soggettiva, dipendente cioè da una valutazione personale dello stato della conoscenza (ovvero, l'espressione di un proprio RATIONAL JUDGEMENT).

4. *Excursus* (o divagazione) - la **sindrome del corvo**

"Il corvo e la gazza paiono capaci di distinguere quantità concrete, dall'uno al quattro. Tobias Dantzig [*Number, the Language of Science*, Macmillan, New York 1959; trad. it. *Il numero, linguaggio della scienza*, La Nuova Italia, Firenze 1965] riferisce in proposito che un castellano aveva deciso di uccidere un corvo che aveva avuto la malaugurata idea di fare il nido nella torre di scolta del castello: più volte egli aveva tentato di sorprendere l'uccello, ma al suo passo questo fuggiva appollaiandosi su un albero vicino, donde tornava non appena l'uomo aveva lasciato la torre. Il castellano ricorse a uno stratagemma: fece entrare due suoi amici nella torre; dopo pochi minuti uno si allontanò, mentre l'altro rimase. Ma il corvo, non cadendo nella trappola, attese che anche il secondo se ne andasse, per riprendere il suo posto. La volta seguente entrarono tre uomini, di cui due si allontanarono, cosicché il terzo potesse attendere quanto voleva per afferrare il corvo; ma l'astuto volatile fu più paziente di lui. Si riprese allora l'esperimento con quattro uomini, ancora senza successo. L'inganno si rivelò conclusivo solo con cinque persone, poiché il nostro corvo non era in grado di distinguere quattro uomini da cinque".
[Citazione da Georges Ifrah, *Storia universale dei numeri*, Mondadori, Milano, 1983, p. 13.

Diremo **sindrome del corvo** quella patologia dell'intelletto a causa della quale si tende a diminuire il numero dei casi da esaminare di qualsiasi problema. Per esempio, dalla complessità 6 per un generico caso di scomparsa, si passa spesso a 3 o addirittura a 2 (fastidiosissimo in particolare il caso in cui, accertata la morte della persona scomparsa, si sente ripetere incessantemente l'interrogativo "omicidio o suicidio?", con cui si riduce a 2 quello che è invece un bel 3). Tale collasso dell'informazione al 2 è abituale presso le persone ottuse (o pigre), e ci piace ricordare nell'occasione il collega Odifreddi il quale si rammarica del frequente ricorso all'esortazione (intimazione) "risponda con un sì o con un no" da parte di certi magistrati nel corso di un processo, quando appunto il "sì" può essere sbagliato allo stesso modo che il "no".
Non possiamo concludere il presente *excursus* senza sottolineare che un grave esempio di sindrome del corvo è riscontrabile nella tendenza filosofica "moderna" che si battezza con il nome di "riduzionismo", ovvero nel passaggio addirittura dal due all'uno. Ecco qui di seguito le tappe più rilevanti di un percorso che cominciò certo bene, ma si sviluppò successivamente sempre peggio (eterogenesi), al punto da far pensare ad un'unica complessa regia dietro il suo sviluppo.
1 - ecumene vs "Nuovo Mondo" (ovvero, la "rivoluzione geografica", Tolomeo vs Colombo, Vespucci, etc.)
[E' questa la PRIMA tappa del lungo cammino riduzionista, coincidente con il "vero" inizio della cosiddetta "rivoluzione scientifica", circostanza questa oggi compresa da pochi... Per maggiori dettagli:

http://www.cartesio-episteme.net/indice1.htm]
2 - terra/cielo (ovvero, la cosiddetta "rivoluzione astronomica", Aristotele vs Copernico-Galileo)
3 - materia/spirito (da dopo Cartesio in poi; fu Cartesio secondo Cornelio Fabro il primo responsabile di questa evoluzione del riduzionismo, sostenendo che tutta la filosofia moderna, appunto da Cartesio in avanti, tende all'ateismo: "il pensiero moderno e' essenzialmente ateo, perche' fondato sul principio d'immanenza, fin da principio", "Introduzione all'ateismo moderno", Studium Ed., Roma, II edizione riveduta, 2 voll., 1969, vol. 1, p. 80. Non sono d'accordo con codesta interpretazione del pensiero cartesiano da parte dell'illustre filosofo, con il quale ho avuto l'onore di parlare diverse volte nella struttura religiosa romana dove si era ritirato a trascorrere gli ultimi anni della sua vita.)
4 - uomo/animale (Darwin)
[E' questo il primo e fondamentale punto di arrivo, ma non l'epilogo, del cammino filosofico che stiamo tracciando, ossia un'opinione dalla quale dipendono tutte le tappe successive. "Il dogma dell'uomo moderno si riduce alle tre umiliazioni dell'umanità constatate da Sigmund Freud: la decentralizzazione della terra, l'interpretazione della vita come un fenomeno di vecchiaia della natura, la riduzione dell'ego umano ad un centro di aggressione e di sessualità" (Erich Kaehler, "Il regno delle idee", Atti del Convegno Internazionale di Geometria a Celebrazione del Centenario della Nascita di Federigo Enriques, Milano, 1971, p. 153). Detta affermazione da parte di un grande matematico fa riferimento al contenuto di un articolo di Freud apparso nel 1917 sulla rivista viennese *Imago* con il titolo "Una difficoltà della psicoanalisi": "Vorrei mostrare come al narcisismo universale, all'amor proprio dell'umanità, siano state fino ad ora inferte **tre gravi umiliazioni** da parte dell'indagine scientifica. a) Dapprima, all'inizio delle sue indagini, l'uomo riteneva che la sua sede, la terra, se ne stesse immobile al centro dell'universo, mentre il sole, la luna e i pianeti si muovevano attorno ad essa con traiettorie circolari. [...] La posizione centrale della terra era comunque una garanzia per il ruolo dominante che egli esercitava nell'universo, e gli appariva ben concordare con la sua propensione a sentirsi il signore del mondo. La distruzione di questa illusione narcisistica si collega per noi al nome e all'opera di Niccolò Copernico nel sedicesimo secolo. [...] Quando tuttavia essa fu universalmente riconosciuta, l'amor proprio umano subì la sua **prima** umiliazione, quella **cosmologica**. b) L'uomo, nel corso della sua evoluzione civile, si eresse a signore delle altre creature del mondo animale. Non contento di tale predominio, cominciò a porre un abisso fra il loro e il proprio essere. Disconobbe ad esse la ragione e si attribuì un'anima immortale, appellandosi a un'alta origine divina che gli consentiva di spezzare i suoi legami col mondo animale. [...] Sappiamo che le ricerche di Charles Darwin e dei suoi collaboratori e predecessori hanno posto fine, poco più di mezzo secolo fa, a questa presunzione dell'uomo. L'uomo nulla più è, e nulla di meglio, dell'animale; proviene egli stesso dalla serie animale ed è imparentato a qualche specie animale di più e a qualche altra di meno. Le sue successive acquisizioni non consentono di cancellare le testimonianze di una parità che è data tanto nella sua struttura corporea, quanto nella sua disposizione psichica. E questa è la **seconda** umiliazione inferta al narcisismo umano, quella biologica. c) La terza umiliazione, di

natura psicologica, colpisce probabilmente nel punto più sensibile. L'uomo, anche se degradato al di fuori, si sente sovrano nella propria psiche. [...] le due spiegazioni - che la vita pulsionale della sessualità non si può domare completamente in noi, e che i processi psichici sono per sé stessi inconsci e soltanto attraverso una percezione incompleta e inattendibile divengono accessibili all'Io e gli si sottomettono - equivalgono all'asserzione che l'Io non è padrone in casa propria. Esse costituiscono insieme la **terza** umiliazione inferta all'amor proprio umano, quella che chiamerei **psicologica**".]

5 - geometria/aritmetica (Karl Weierstrass & C. - costruzione dei numeri reali su base puramente aritmetica)

6 - spazio/tempo (Albert Einstein, Hermann Minkowski: "Le vedute sullo spazio e sul tempo che desidero esporre davanti a voi sono sorte dal terreno della fisica sperimentale, ed in esso risiede la loro forza. Queste vedute sono radicali. D'ora in poi lo spazio preso a sé stante, e il tempo preso a sé stante, sono condannati a scomparire come pure ombre, e soltanto una sorta di unione dei due conserverà una propria realtà indipendente", da una lezione a Colonia tenutasi il 21 settembre 1908, successivamente pubblicata con il titolo "Raum und Zeit", *Physikalische Zeitschrift*, 10, 104, 1909.)

7 - determinismo/indeterminismo (fisica quantistica, Pierre Simon Laplace vs Werner Heisenberg, fino al famoso paradosso detto del "gatto di Schroedinger" - Se nell'800 si tento' senza successo di "determinare lo spirito" (concezione peraltro ancor oggi diffusa presso molti scienziati), nel '900 si cercò inversamente - con il medesimo fine "riduzionista"! - di "indeterminare la materia" (proponendo di assimilare quindi il "libero arbitrio" e la coscienza allo stesso tipo di indeterminazione di una roulette), ma il fallimento dell'operazione e' stato identico.)

8 - uomo/macchina (Turing - qualsiasi "registratore" ricorda le cose meglio di un essere umano, ma nessun tipo di intelligenza artificiale potrà mai sostituire, se non in apparenza, la capacità di "scelta", appunto **intelligenza**, che dipende dalla *voluntas*.)

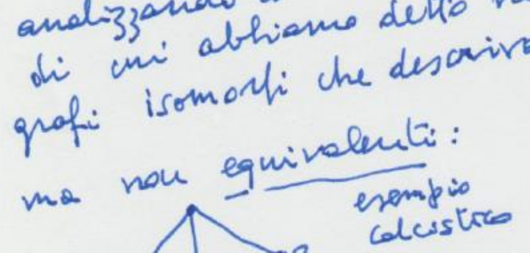
9 - maschio/femmina (la "cultura" LGBT etc. del tempo presente).

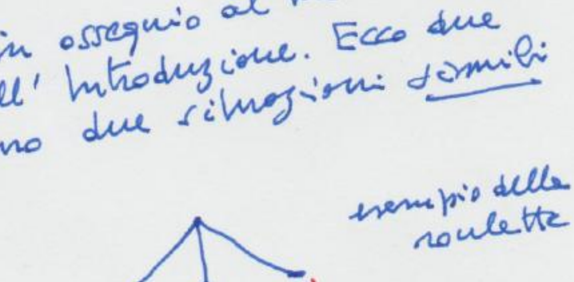
(un simile rilievo comprende anche il tentativo, **umanistico e non scientifico**, di eliminare il concetto di "razza" umana, vedi per esempio Odifreddi: https://www.youtube.com/watch?v=-cRyCe9Tt4c).

5. L'emergenza dello 0 e dell'1

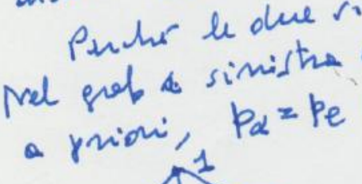
È questa una questione che riteniamo importante per il nostro approccio alla logica della probabilità, anche se all'atto pratico il problema si potrebbe ovviare sostituendo a 0 un $1/n$ per $n \gg 0$, e a 1 un $1-1/n$ o un analogo n, ovvero un $1-\varepsilon$ con $0<\varepsilon \ll 1$, preferiamo però continuare a rimanere in ambiti chiusi anziché aperti; ciò che risulterà particolarmente utile al momento della definizione di quantità d'informazione (cap. 6).

Prendiamo la cosa da lontano, cominciando a studiare i problemi posti dal passaggio da probabilità a priori (relative ed assolute) a probabilità a posteriori (ancora relative ed assolute). Faremo ciò analizzando due esempi, in ossequio al metodo ostensivo di cui abbiamo detto nell'Introduzione. Ecco due grafi isomorfi che descrivono due situazioni simili ma non equivalenti:

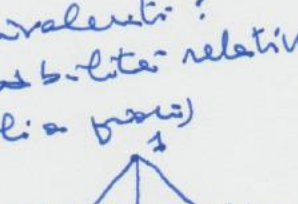




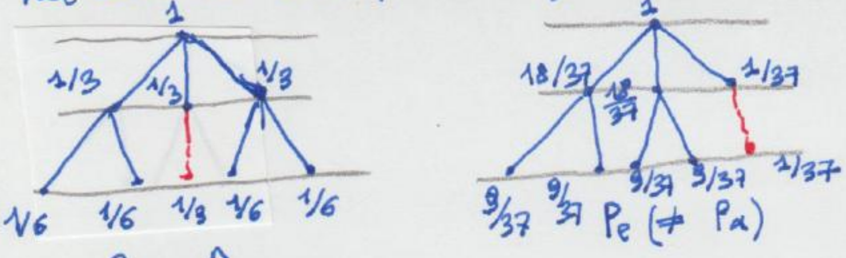
Perché le due situazioni non sono equivalenti? Nel grafo a sinistra abbiamo le seguenti probabilità relative a priori, $p_a = p_e$ (alternative [illegible] equiprobabili a priori)



in quello di dx le p_a sono ≠ dalle p_e e precisamente: queste



Queste invece le corrispondenti probabilità assoluta (21)
nel caso dei corrispondenti grafi compattificati: (sempre in rosso)



$P_a = P_e$ $\qquad$ $P_e \ (\neq P_a)$

Questo ci basta per asserire che le due situazioni sono simili ma non equivalenti. Cosa accade (può accadere) però se passiamo da probabilità a priori a probabilità a posteriori? Ecco che, come per incanto, potremmo trovarci di fronte a due grafi non solo isomorfi, ma con un isomorfismo che preserva le probabilità dei nodi corrispondenti, ovvero una situazione di apparente equivalenza. Solo apparente, ribadiamo perché contingente ad un dato osservatore, ed al suo stato della conoscenza delle situazioni ad un dato istante t, quindi ad una caratteristica che può cambiare sia al variare dell'osservatore a, sia al variare del tempo t.

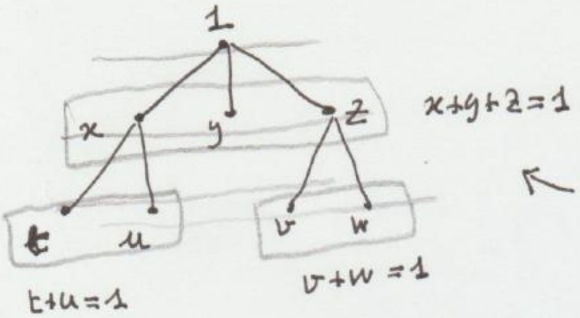
Ciò premesso, ci pare adesso opportuno cominciare a chiedersi se un passaggio da una P_e a una p (caso assoluto), o da una P_e a una P (caso relativo), sono passaggi assolutamente arbitrari, oppure soddisfare alcune condizioni di logica e di coerenza. È qui che la questione dell'emergenza dello 0 o dell'1 comincerà a farsi rilevante. Bene, dato un certo grafo (compatto o non compatto) Γ descrivente una certa situazione, abbiamo da esaminare prima di tutto le sue 4 probabilità a priori: $p_a:\Gamma\to I$, $P_a:\Gamma\to I$, $p_e:\Gamma\to I$, $P_e:\Gamma\to I$ e ci rendiamo subito conto che qui non possono mai apparire né lo 0 né l'1, a meno che non si tratti di una probabilità relativa ad un ramo solitario rosso (che può apparire in una compattificazione).

In tutti gli altri casi un 1 non ~~può~~ potrà darsi
senza la contemporanea presenza di uno 0
allo stesso livello dell'1, e quindi è sull'emergenza
dello 0 che concentreremo la nostra attenzione.
Abbiamo detto che 0 non potrà mai apparire né per
pa né per P_a, e ciò per la stessa definizione pascaliana
di probabilità. Ma lo stesso varrà anche per p_e e P_e.
Se infatti uno 0 comparisse tra le probabilità a priori
effettive (relative o assolute), ciò starebbe semplicemente
a significare che c'è stato un errore nella costruzione
del grafo Γ descrivente la situazione sotto esame. Il
grafo avrebbe cioè dovuto essere semplificato omettendo
appunto quella data modalità con lo 0, in quanto
ritenuta effettivamente impossibile. Tale
procedimento di cancellazione, ovvero di ridefinizione
del grafo associato alla situazione, dovrà necessariamente
tenere conto anche di tutte le relative conseguenze logiche,
ma su ciò ci sembra ormai di poter sorvolare, e
tornare alla nostra questione principale, l'emergenza
dello 0. Uno 0 potrà apparire, per quanto
precedentemente detto, solo per le probabilità a
posteriori $p: \Gamma \to I$, $P: \Gamma \to I$, in conseguenza
di una variazione dello stato dell'informazione
dell'osservatore ω dal tempo t_0 al tempo t:

$$\mathcal{I}(\omega, \mathcal{J}, t_0) \Rightarrow \mathcal{I}(\omega, \mathcal{J}, t) \quad \left(\text{ricordiamo che } p = p(t),\ P = P(t)\right).$$

Vediamo adesso con qualche esempio come ciò possa
accadere. [VEDI IN PROPOSITO ANCHE L'APPENDICE CHE
ABBIAMO AGGIUNTO A QUESTA PRIMA PARTE
DELLA NOSTRA INVESTIGAZIONE]

NOTA A LUNGO RIMANDATA: Preferiamo l'approccio proposto a quello relativo ad una possibile algebra degli eventi, che contempli i soliti connettivi logici ∧, ∨, ¬ (negazione), che costringerebbe ad una valutazione di indipendenza oppure no di modalità che appaiono in grafi diversi. È invece lecito introdurre $\mu \wedge \mu'$, $\mu \vee \mu'$, $\neg\mu$, per singole modalità dello stesso livello, con $P(\mu \vee \mu') = P(\mu) + P(\mu')$, $P(\neg\mu) = 1 - P(\mu)$, $P(\mu \wedge \mu') = P(\mu)P(\mu')$.

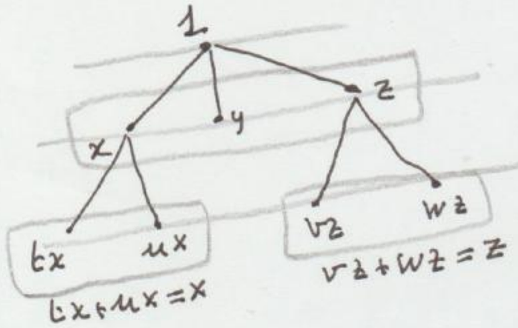


$p(t_0) = p_e$
$P(t_0) = P_e$

ma anche si può passare da $p(t), P(t)$ a $p(t'), P(t')$

a posteriori relativa $p(t_0)$
su Γ

3 sezioni, 1 al 1° livello, 2 al 2°
la Σ delle probabilità su ciascun livello è uguale ad 1.

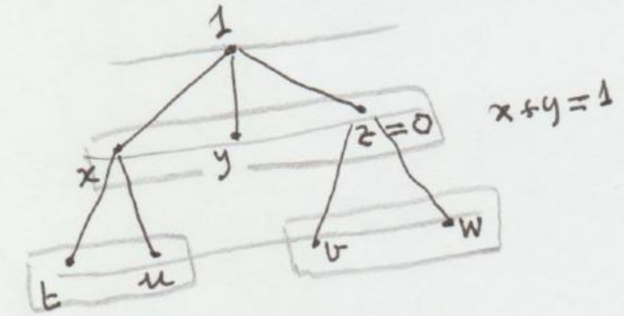


a posteriori ~~relative~~ assolute $P(t_0)$
su Γ

3 sezioni come prima, la Σ delle probabilità è uguale alla probabilità della relativa radice.

su ∀ sezione

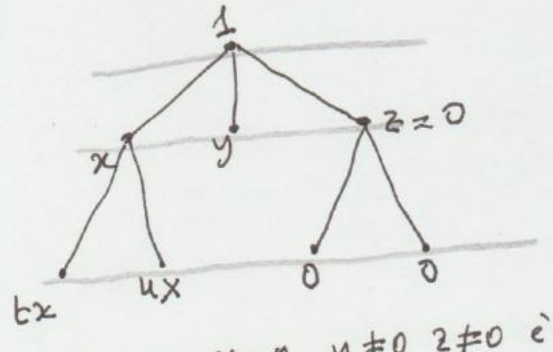
$p(t)$



$x \neq 0$ $y \neq 0$

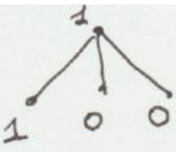
Tutto rimane invariato nonostante la comparsa ~~[illegible]~~ 0 al 1° livello, di uno non così per le probabilità assolute.

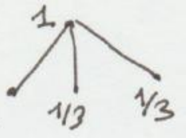
Ovvero, la comparsa di quello 0 al 1° livello ha implicato altri due 0 al 2° livello, e resa impossibile la ricostruzione della $p(t)$ dalla $P(t)$.



Il caso $x=0$ $y \neq 0$ $z \neq 0$ è analogo.

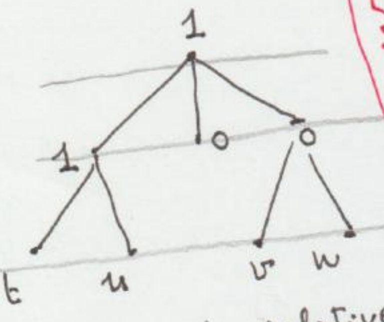
Nel caso $y=0$ (ma $z \neq 0$) tutto è ancora più facile. Nel caso $y=z=0$ e quindi $x=1$ siamo di fronte ad una certezza al 1° livello.

Tricotomia al 1° livello  certezze

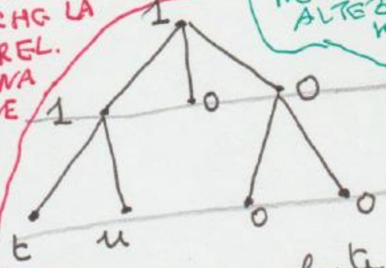
(diremo per contro  l'incertezza assoluta, ci torneremo sopra nel prossimo capitolo)

TUTTO DIVENTA OVVIAMENTE PIÙ SEMPLICE NEL CASO DI GRAFI DI ALTEZZA 1 OVVERO K-TOMIE, PER LE QUALI RIMANE LA SOLA CONDIZIONE $\sum p(\mu)=1$

INSOMMA, BISOGNA RISPETTARE LE CONDIZIONI CHE LA Σ DELLE PROB. REL. DEI NODI DI UNA STESSA SEZIONE DEVE ESSER UGUALE AD 1, CONDIZIONE CHE VARRÀ PER LA Σ DELLE PROB. ASS. DEI NODI DI UN STESSO LIVELLO



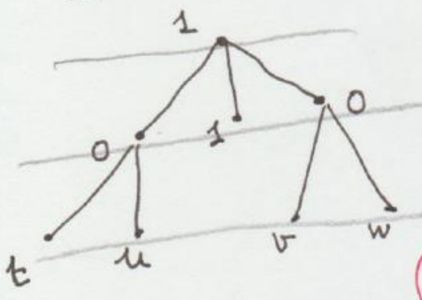
prob. relative



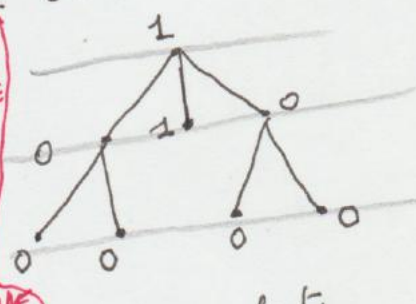
prob. assolute

Se invece $x = 0$ insieme a $z = 0$ abbiamo ovviamente:

(PER LE PROB. ASSOLUTE DOVRÀ ANCHE VALERE LA CONDIZIONE CHE LA Σ DELLE PROB. DEI NODI DI UNA STESSA SEZIONE È UGUALE ALLA PROB. DELLA RADICE (ASS.) DELLA SEZIONE.



prob. relative

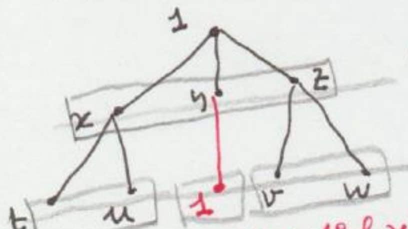


prob. assolute

Facile discutere adesso il caso Γ_c, grafo compattificato.

ANCHE QUI LO 0 NON POTRÀ MAI APPARIRE PER LE 4 PROB.TÀ A PRIORI

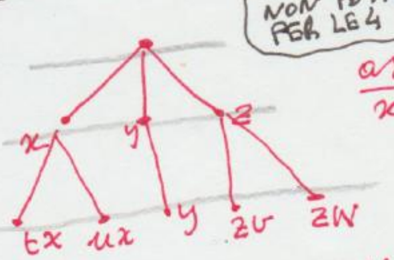
relative



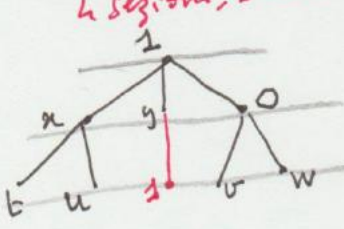
4 sezioni, 1 al 1° livello, 3 al 3°

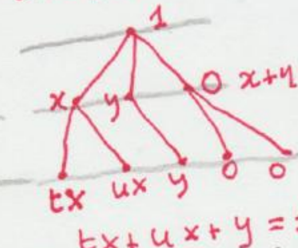
assolute

$x+y+z=1$



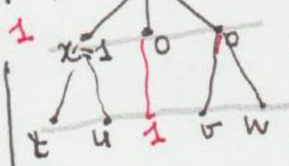
$tx+ux+y+zv+zw=1$

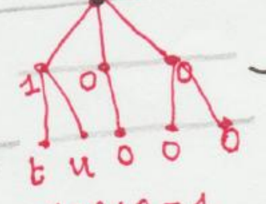




$x+y=1$

$tx+ux+y=1$





$1+0+0=1$

$t+u+0+0+0=1$

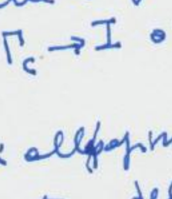
relative vs assolute, caso di un solo 0, caso di due 0.

Si noti: esplicitamente che una certezza ad un dato livello non implica necessariamente l'esistenza di ulteriori certezze in altri livelli.

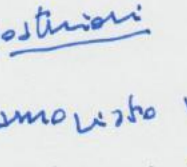
Chiudiamo il capitolo ritornando sul caso dell'1 (sempre radice a parte), che è evidentemente collegato a quello dello 0 in virtù della circostanza che varie volte abbiamo a che fare con probabilità la cui somma deve necessariamente essere 1. Tanto per riepilogare, l'1 non potrà mai apparire nelle le probabilità a ~~priori~~ di dominio Γ, e nemmeno nelle probabilità assolute a priori di dominio Γ_c, $P_a : \Gamma_c \to I$ e $P_e : \Gamma_c \to I$, ma soltanto nelle Σ probabilità a priori relative di dominio Γ_c, in corrispondenza a quelli che abbiamo detto "rami solitari" (ovvero, a sezioni di Γ_c di cardinalità uguale ad 1). Insomma, una comparsa significativa dell'1 potrà darsi soltanto nelle le probabilità a posteriori $p : \Gamma \to I$, $p : \Gamma_c \to I$ (relative) oppure $P : \Gamma \to I$, $P : \Gamma_c \to I$ (assolute)

ASSOLUTE — È INSOMMA SULLE P CHE BISOGNERÀ AGIRE, NON SULLE p RELATIVE!

Abbiamo visto negli esempi di cui alle pagine precedenti alcuni casi di comparsa significativa dell'1, ma vogliamo adesso ricapitolare il tutto analizzando alcuni esempi particolari. Partiamo per esempio dal solito grafo

Γ: 

, completiamolo

Γ_c: 

e consideriamo qui le probabilità a posteriori $p(t) : \Gamma_c \to I$, $P(t) : \Gamma_c \to I$ (rispettivamente relative ed assolute) dipendenti dallo stato dell'informazione $\mathcal{I}(w, \mathcal{S}, t)$. Se in un passaggio $\mathcal{I}(w, \mathcal{S}, t) \Rightarrow \mathcal{I}(w, \mathcal{S}, t')$ vogliamo modificare la nostra analisi probabilistica, come potremo fare? con logica e coerenza. Possiamo semplicemente considerare le sole probabilità assolute, procedendo quindi (quasi) ad arbitrio anche con lo 0 e con l'1, tenendo conto però che su ogni cammino ordinato del tipo ① → ④ → ⑧ le P devono diminuire dall'alto verso il basso, che la Σ delle P ad ogni livello deve fare 1, e che la Σ delle P in ogni sezione deve coincidere con la P della radice della sezione. Rispettando queste condizioni si possono mettere 0 e 1 dove si vuole, alla fine non sarà possibile ricostruire dalla P(t) le corrispondenti p(t), di fronte a forme indeterminate del tipo 0/0, vedi anche l'APPENDICE

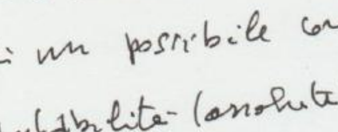
6. Il confronto tra due funzioni di probabilità assolute sul medesimo grafo compatto, e la definizione di quantità d'informazione e di divergenza

Intendiamo occuparci da ultimo del problema di un possibile confronto tra due distribuzioni di probabilità (assolute) $P : \Gamma_c \to I$, $P' : \Gamma_c \to I$
$P = P(t)$ $\quad P' = P(t')$ $\quad t < t'$

possibilmente differenti a seguito di una variazione dello stato dell'informazione $\mathcal{I}(\omega, \mathcal{S}, t) \Rightarrow \mathcal{I}(\omega, \mathcal{S}, t')$.

Esaminiamo da vicino il caso di una k-tomia

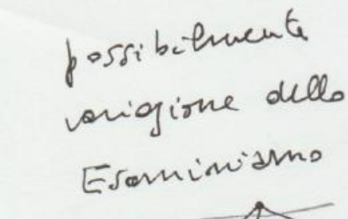
cioè h = altezza = 1
e K = complessità (dell'unico livello)



$P(\textcircled{1}) = P(1) = x_1$ $\quad P(2) = x_2$ $\quad P(K) = x_K$

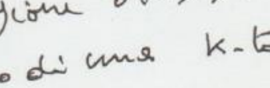
UN K CHE DIVENTERÀ PRESTO UN n PER INVETERATA ABITUDINE DI GEOMETRA!

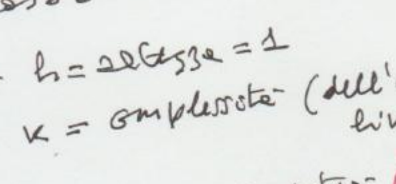
Tenendo conto che un qualsiasi grafo compatto potrà "decomporsi" in h K-tomie, $K = K_1, \dots K_h$:



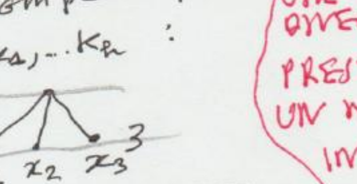
grafo completo $h = 2$ di complessità ordinaria $(3,3)$

$9 = 3+3+3$





[la ricostruzione univoca da destra a sinistra non sarà più possibile se per esempio $x_{13} = 0$:



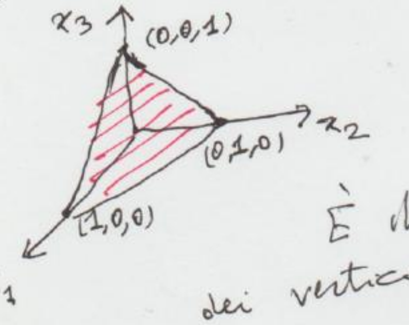
$9 = 2+4+3$)

e potremo confrontare le due probabilità P e P' procedendo livello per livello.

Bene, è chiaro che la funzione P potrà essere identificata con un punto, che denoteremo ancora con $P \equiv (x_1, x_2, \dots, x_n)$, dell'ipercubo $I^n \subset \mathbb{R}^n$, contenuto nell'iperpiano

$$x_1 + x_2 + \dots + x_n = 1 .$$

Risulta quindi che l'"universo" ($\mathcal{U}$) nell'ambito del quale cercheremo di operare un confronto tra due punti P, P' è l'$(n-1)$-simplesso costituito da detta intersezione. Offriamo uno schema della situazione nel caso $n=3$, onde guidare la nostra intuizione geometrica:



Ossia, $\mathcal{U}$ sarà in questo caso il triangolo equilatero indicato in rosso.

È chiaro che potremo parlare dei vertici di $\mathcal{U}$, $\begin{matrix}(1\,0\,\ldots\,0)\\(0\,1\,\ldots\,0)\\(0\,0\,\ldots\,1)\end{matrix}$ come $\begin{matrix}e_1\\e_2\\e_n\end{matrix}$ delle nostre n certezze, e del centro del simplesso $(1/n\ 1/n\ \ldots\ 1/n)$ C come del punto di massima incertezza. Come nel caso del triangolo equilatero, C sarà il centro di un'ipersfera circoscritta ad $\mathcal{U}$, il cui raggio diremo R, e di un'ipersfera inscritta il cui raggio diremo ρ. Semplici calcoli conducono alle seguenti formule (a destra il caso del triangolo equilatero)

Il diametro D del simplesso è sempre uguale a $\sqrt{2}$ — distanza massima tra due punti di $\mathcal{U}$

$D=\sqrt{2}$

$D =$ lato L del triangolo $=\sqrt{2}$

$$R=\sqrt{\frac{n-1}{n}}=\frac{n-1}{n}H=(n-1)\rho \qquad R=\frac{\sqrt{3}}{3}L=\sqrt{\frac{2}{3}}$$

$$\rho=\sqrt{\frac{1}{n(n-1)}}=\frac{1}{n}H=\frac{1}{n-1}R \qquad \rho=\frac{\sqrt{3}}{6}L=\frac{1}{\sqrt{6}}=\sqrt{\frac{1}{6}}$$

$$H=\text{altezza di }\mathcal{U}=n\sqrt{\frac{1}{n(n-1)}}=\sqrt{\frac{n}{n-1}} \qquad H=\frac{\sqrt{3}}{2}L=\sqrt{\frac{3}{2}}$$

$$R\,H=1 \qquad \frac{\rho}{R}=\frac{1}{n-1} \qquad R+\rho=H=\frac{n-1}{n}H+\frac{1}{n}H$$

Tornando alla nostra questione, diremo che la <u>quantità d'informazione</u> necessaria a $\Theta(\omega, S, t)$ per passare dalla <u>massima incertezza</u> $C \equiv (1/n, \dots, 1/n)$ a $P \equiv (x_1, \dots, x_n)$ è misurata da un'espressione del tipo:

$$QI(P) = d(P,C)^2/R^2 = \frac{n}{n-1}\left[(x_1 - \frac{1}{n})^2 + \dots + (x_n - \frac{1}{n})^2\right]$$

quantità d'informazione di P

dove i quadrati vengono introdotti allo scopo di rimanere per quanto possibile nell'ambito <u>razionale</u>, e mettere inoltre le basi per una gradita <u>linearità</u> della situazione.

È chiaro infatti che potremmo introdurre il concetto di quantità d'informazione di P relativa ad una modalità μ con l'espressione: $QI(P|\mu) = \frac{n}{n-1}\left[P(\mu) - \frac{1}{n}\right]^2$, sicché avremo poi ovviamente: $QI(P) = \sum_{\mu} QI(P|\mu)$.

Detto che l'informazione è naturalmente ciò che riduce l'incertezza, osserviamo esplicitamente che:

$0 \leq QI(P) \leq 1$ (ovvero, $QI(P)$ è una <u>percentuale</u>)

$QI(P) = 0$ sse $P = C$, $QI(P) = 1$ sse P = certezza.

Ecco che ci sembra che questa QI sia più facilmente interpretabile del $-\log_2(P(\mu))$ di Shannon, ossia <u>adeguata</u> alla fenomenologia che si intende descrivere.

<u>Nota 1</u> - Se volessimo occuparci più in generale di P (a posteriori assolute) su un qualsiasi grafo compatto Γ_c, potremmo introdurre diverse QI di P ai vari livelli di Γ_c, e parlare (forse) di una QI <u>complessiva</u> come della media aritmetica tra di queste.

VEDI ANCHE PAG. 31

<u>Nota 2</u> - Ci siamo finqui occupati del solo caso $P_\alpha \equiv P_e$, ossia del caso di una equiprobabilità a priori, in cui la massima incertezza è appunto $C \equiv (1/n, \dots, 1/n)$. Se $P_\alpha \neq P_e$, dovremmo scrivere per esempio $QI(P|\mu) = C(P_e)\left[P(\mu) - P_e(\mu)\right]^2$, dove $C(P_e)$ è un opportuno coefficiente dipendente da P_e che va a prendere il posto di $n/n-1$. È chiaro dalla nostra impostazione che $C(P_e)$ proverrà dalla considerazione di $d(P,P_e)^2/d_{max}(P_e)^2$ dove al denominatore appare la max distanza di P_e da una certezza.

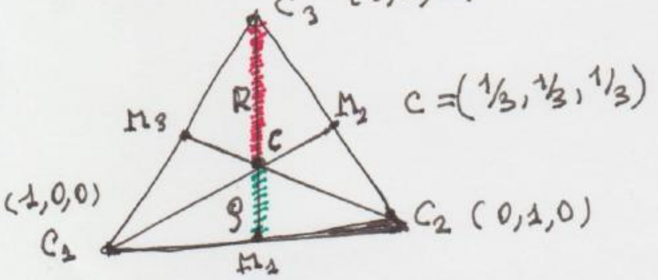
Nota 3 - Ci fa notare un caro ex studente, GDS, che il senso della definizione di Shannon si riferisce al sentimento di sorpresa che per ω costituisce il verificarsi dell'evento $\mathcal{E}$ nella modalità μ (data ovviamente la nostra previsione soggettiva $P(\mu)$). Ovvero, se si verifica una modalità che per noi era certa $(P(\mu)=1)$ il senso di sorpresa - i.e., la quantità dell'informazione recata dall'evento secondo Shannon - sarà 0, mentre sarà viceversa ∞ se $\mathcal{E}$ si avvera in una modalità μ che ritenevamo impossibile, cioè $P(\mu)=0$. Tale QI si riferisce quindi ad un momento successivo al verificarsi dell'evento secondo una delle modalità previste, ossia ad un momento in cui non vi è più nessuna incertezza, ed esula quindi a nostro parere dall'ambito della nostra analisi.

Nota 4 - Vale allora la pena ancora di comprendere meglio il significato della $\left(\frac{n}{n-1}\right)\left(p-\frac{1}{n}\right)^2$.

Tale espressione è 0 sse $p=\frac{1}{n}$, ovvero ci troviamo nella massima incertezza in relazione al possibile verificarsi di $\mathcal{E}$ nella modalità μ. Se $p=0$ essa vale $\frac{n}{n-1}\frac{1}{n^2}=\frac{1}{n(n-1)}$, un valore che tende a 0 per $n\to\infty$. Ciò si interpreta pensando che se $n\gg 0$, allora il passaggio dalla probabilità $1/n$ alla probabilità 0 richiede poco sforzo (cioè poca quantità di informazione), comunque non 0 (si pensi al caso dell'estrazione di un biglietto della lotteria con 1000 biglietti venduti). Di contro, per $p=1$ otteniamo $\frac{n}{n-1}\left(\frac{n-1}{n}\right)^2=\frac{n-1}{n}$ il cui limite per

$n \to \infty$ è uguale ad 1. Ovvero, nel caso del passaggio da $(1/n, \dots, 1/n)$ ad una certezza $(1, 0, \dots, 0)$, abbiamo una QI che vale $\frac{n-1}{n}$ per la modalità che diventa 1, e $(n-1)$ volte $\frac{1}{n(n-1)}$ per le $(n-1)$ modalità che da $1/n$ passano a 0, ovvero in totale $\frac{n-1}{n} + (n-1)\frac{1}{n(n-1)} = 1 - \frac{1}{n} + \frac{1}{n} = 1$.

Può diventare adesso istruttivo esaminare la situazione in maniera più approfondita considerando il nostro universo $\mathcal{U}$ bidimensionale, ossia $\mathcal{U}$ = triangolo equilatero, ossia il caso $n=3$, tricotomia,



D = diametro = lato $= \sqrt{2}$

$R = d(C, C_1) = d(C, C_2) = d(C, C_3) = \sqrt{2/3}$

$\rho = d(C, M_1) = d(C, M_2) = d(C, M_3)$ $\quad M_i$ $i = 1,2,3$ punti medi dei lati $\rho = \sqrt{\frac{1}{6}}$

$d(C_1, M_1) = \sqrt{2}/2$ etc.

È chiaro che il nostro universo $\mathcal{U}$ delle probabilità si può ripartire in 5 sottoinsiemi, 1 chiuso e 4 aperti (relativi):

- I "punti" P che hanno QI uguale esattamente a $\frac{\rho^2}{R^2} = \frac{1}{(n-1)^2} = \frac{1}{4} = 25\%$;
- I punti P che hanno QI minore della precedente, ossia minore del 25%, e che sono quelli <u>interno</u> al cerchio inscritto ad $\mathcal{U}$;
- I punti P che hanno QI maggiore del 25%, i quali si possono ripartire nei 3 sottoinsiemi aperti (relativi) per i quali rispettivamente C_1 è la certezza più <u>vicina</u>, oppure C_2, oppure C_3.

Possiamo formalizzare meglio quest'ultima definizione introducendo per due punti P, P' di $\mathcal{U}$ una distanza relativa $\delta(P,P') = \frac{d(P,P')}{D}$.

Allora per $\forall$ punto P' di $\mathcal{U}$ possiamo introdurre le 3 distanze relative $\delta(P, C_i)$ $i = 1,2,3$, ed introdurre i concetti di distanza relativa minima e massima. Gli ultimi 3 sottoinsiemi (aperti relativi) introdotti poc'anzi sono quelli per cui rispettivamente

$$\delta_{min}(P) = \delta(P, C_1), \quad \delta_{min}(P) = \delta(P, C_2), \ldots$$

Ovviamente: $\delta_{min}(P) + \delta_{max}(P) \geq 1$

Qualche esempio:

		δ_1	δ_2	δ_3
C_1	$\Rightarrow$	$\delta_1 = 0$	$\delta_2 = 1$	$\delta_3 = 1$
C_2	$\Rightarrow$	1	0	1
C_3	$\Rightarrow$	0	0	1
C	$\Rightarrow$	R/D	R/D	R/D
M_1	$\Rightarrow$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{R+P}{D} = \sqrt{\frac{3}{2}}$ etc.

$R/D = \sqrt{\frac{1}{3}} = \frac{\sqrt{3}}{3}$

Concludiamo il presente capitolo, e l'intero scritto, introducendo due ulteriori concetti che ci sembrano significativi.

Abbiamo definito per $\forall\, P \in \mathcal{U}$ una sua $\mathcal{Q}\,I(P)$, $P(t): \Gamma_C \to I$ (assoluta) e se interpretiamo P come una $P(t): \Gamma_C \to I$ (assoluta) ecco che possiamo volerla confrontare con una $\mathcal{Q}\,I(P')$ ove $P' = P(t'): \Gamma_C \to I$ per qualche $t' > t$.

Stiamo quindi parlando del passaggio da una $\mathcal{I}(\omega, \mathcal{S}, t)$ che determina P ad una $\mathcal{I}(\omega, \mathcal{S}, t')$ che determina P', passaggio che può avvenire con l'affluenza ad $\mathcal{I}(\omega, \mathcal{S}, t)$ di una singola notizia ».

La prima cosa che può venirci in mente è quella di introdurre come <u>misura della variazione</u> (o divergenza) il semplice incremento

$$\Delta QI = QI(P') - QI(P).$$

Questo sarà un valore uguale a 0 ses P', P hanno la stessa quantità d'informazione, negativo se la QI è diminuita (ovvero, se nel passaggio da P a P' ci siamo avvicinati a C ed allontanati da qualche certezza), positivo in caso contrario (nel passaggio da P a P' ci siamo allontanati da C, ed avvicinati quindi a qualche certezza).

ossia la stessa distanza da C

Δ potrà però essere 0 anche quando P e P' sono assai diverse, per esempio nel caso di un <u>ripensamento</u>, quando si passa cioè per effetto della notizia ν da una certezza C_i ad una delle certezze "opposte" C_j, $j \neq i$. (in entrambi i casi abbiamo in effetti una stessa QI massima possibile).

Appare allora opportuno per la misura in oggetto (ovvero, del concetto di <u>peso</u> della notizia ν) introdurre la semplice distanza relativa

$$\delta(P,P') = \frac{d(P,P')}{D} \quad (\text{o se si preferisce } \delta^2)$$

che è in effetti un numero compreso tra 0 e 1, uguale a 0 ses $P=P'$, ed $=1$ ses P e P' sono entrambe certezze (certezze opposte).

Risulta chiaro da quanto precede che:

$$\text{divergenza } \operatorname{div}(P,P') = \delta(P,P')^2 = \frac{d(P,P')^2}{D^2}$$ risulta uguale, ponendo $P=C$ e $P'=P$, a:

$$\operatorname{div}(C,P) = \frac{d(C,P)^2}{D^2} = \frac{R^2\, QI(P)}{D^2} = \frac{(n-1)}{2n}\, QI(P).$$

FINE

Dedicato a GDS e a Francia (Sant'Ippe)

INDICE

UB, terminato di scrivere a Perugia, il 7 agosto 2023

- - - - -

P.S. Non possiamo dimenticare di citare il curatore del seguente sito (Giovanni Fausti), un cui sincero commento ha stimolato il nostro lavoro:

https://webcrew.it/teoria-informazione/#:~:text=Non%20avventurandosi%20pi%C3%B9%20di%20tanto,'informazione%20che%20esso%20trasmette%E2%80%9D

> Non avventurandosi più di tanto nella ricostruzione teorica dell'articolo di Shannon, il Dizionario di filosofia sintetizza: "In generale si può dire che più un messaggio è improbabile, maggiore è l'informazione che esso trasmette". **Sarebbe interessante capire più nel dettaglio come la teoria dell'informazione sia arrivata a questa affermazione, anzi se siete esperti in materia, spiegatecelo nei commenti.** [...] Questa ricostruzione, molto approssimativa e parziale, può comunque essere utile, credo. Una cosa che mi ha colpito, lo scrivevo sopra, è di come tale teoria sia nota come Teoria dell'informazione, eppure Shannon nel titolo del suo importante articolo scrive "communication" e non "information". Quest'altra parola, "comunicazione", riguarda anch'essa questo blog se non altro perché è la parola in cui si è soliti racchiudere il settore in cui lavoriamo. Comunicazione e informazione sono sinonimi? Sono parole intercambiabili?