PROBABILITÀ E ... ALBERI DI NATALE,
CON UN'APPLICAZIONE AL PARADOSSO DI MONTY HALL

Cap. 0 - Preambolo

In tempi recenti ho dedicato parecchie riflessioni al modo migliore di esporre il cosiddetto paradosso di Monty Hall (o delle 3 porte) anche a persone non particolarmente abituate al linguaggio matematico
(vedi: http://www.cartesio-episteme.net/probabilita.htm).
Ultimamente ho deciso invece di fare un ricorso più preciso a detto linguaggio, ritenendo che i suoi conseguenti "automatismi" (**algoritmi**) potessero essere utili a chiarire ogni aspetto del problema (ovvero, a facilitarne la comprensione), e tenuto conto che si tratta di una matematica assai elementare, vale a dire di livello decisamente scolastico, alla peggio liceale. L'unica difficoltà scaturisce (potrebbe scaturire) dalla non abitudine al "linguaggio" matematico, ossia, **nomenclatura** e **simbolismo** (semplice "stenografia" matematica), al quale basta del resto abituarsi (in caso sopportando i relativi duplici "abusi", **di notazione** e **di linguaggio**, a volte inevitabili se non si vuole peccare di eccessiva pignoleria, o pesantezza formale - ricordo una collega assistente a Roma che mi chiamava "cacaspilli"), ovviamente ... volendo. Volontà che dovrebbe essere peraltro desiderio di tutti, trattandosi in simili contesti soltanto di **logica** e non di **matematica** (men che meno di **logica matematica**, termine che tecnicamente indica tutt'altra materia peraltro abbastanza difficile, vedi per esempio: http://www.cartesio-episteme.net/mat/teor-goed.pdf). Insomma, ci troveremo di fronte ad una semplice descrizione delle **leggi dell'intelletto** (in pratica effettuata esclusivamente attraverso l'introduzione di **insiemi** e **funzioni**), leggi che debbono essere utilizzate anche quando si vuole dare una descrizione abbastanza **completa** di tutte le diverse possibilità di ... una ricetta gastronomica alquanto "semplice" come la famosa carbonara: guanciale o pancetta? quante uova per commensale? solo tuorli o anche albumi? pecorino o parmigiano? *etc.*.
Da tale proposito sono nate alcune pagine che ho dedicato al mio caro figlio Giacomo nel giorno del suo 38mo compleanno (avendo ricordato che fece cenno alla questione nel suo video: https://www.youtube.com/watch?v=OryF_6yPL4o - *Quite Strange for an Octopus*), pagine che consideravo alquanto decenti. Di contro a tale opinione, ho ricevuto invece diversi rilievi fondati principalmente su una certa ambiguità del simbolismo ivi impiegato*, i quali rilievi, uniti ad una preesistente analoga insoddisfazione nei confronti della nomenclatura e del simbolismo da me utilizzati nel corso delle suddette riflessioni, mi hanno spinto alla presente profonda revisione del citato anomalo "regalo di compleanno". Da codesta finalità sono venute fuori le pagine che seguono (la cui lettura sarà probabilmente riservata unicamente al nominato figlio, e a qualche ultimo caro

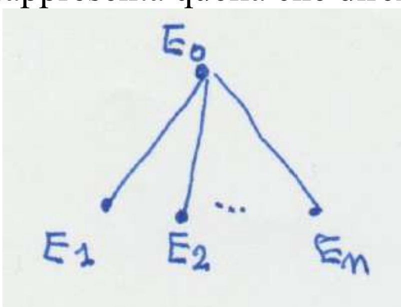
amico), una sorta di materiale preparatorio direi per una serie di lezioni liceali sulla probabilità, prendendo come spunto una discussione del famoso paradosso. Detto che ho scelto un'esposizione per quanto possibile sintetica, e sotto il profilo metodologico più **ostensiva** (vale a dire, definizioni implicitamente introdotte attraverso esempi concreti) che non **intensiva** (vale a dire, definizioni esplicite "nominalistiche"), e che probabilmente non sono riuscito a non dare l'impressione di voler tenere i piedi in due staffe, senza peraltro farlo bene con nessuna delle due, non mi resta che augurare buona lettura, nella speranza di essere stato comunque utile a qualcuno.

* Una questione che conduce la mia mente ad una riflessione epistemologico-didattica. Uno degli stimatissimi professori della mia giovinezza (comunque, altri tempi, quelli della tanto deprecata università dei "baroni", docenti i quali ancora sapevano qualcosa della loro materia, e non si limitavano a cercare di fare i "simpatici" con gli studenti, promovendo tutti con buoni voti, parlando loro di politica o di attualità, limitandosi a fare lezione mostrando sempre gli stessi lucidi anno dopo anno, a tirare insomma "quattro paghe per il lesso"**) diceva sempre che la matematica è soprattutto "economia di pensiero". Dopo tanti anni di insegnamento, ritengo però oggi che, a fare eccessiva economia (quella tra l'altro costellata di vaghi incisi del tipo "Non è restrittivo supporre", "Senza perdere in generalità", "Introduciamo un abuso di notazione - o di linguaggio - tanto ci si capisce lo stesso", *etc.*), si rischia di indurre infine in confusione. Ovvero, è forse preferibile essere alquanto accurati all'inizio dell'esposizione di un problema, e solo in seguito economizzare-semplificare. La strada inversa è assai più ardua e spesso non viene affatto percorsa. Ricordo quei colleghi che nei primi anni degli studi di matematica (o di fisica) spiegavano un dato argomento in maniera piuttosto raffazzonata, dicendo agli studenti che lo avrebbero rivisto in maniera più esatta in altri corsi degli anni successivi, evidentemente ignorando, o fingendo di ignorare, che molti dei giovani cui si rivolgeva quell'informazione detti corsi non li avrebbero mai seguiti, e l'argomento non l'avrebbero quindi mai più rivisto per bene.

** In relazione a tale dilagante fenomeno, che mi spinse nell'ormai lontano anno 2005 ad abbandonare per sempre l'amato insegnamento universitario (al quale avevo dedicato l'intera vita), stanco di vedere la proliferazione dei "mercanti nel tempio" senza poter fare nulla per contrastarla (qui un esempio di una lotta purtroppo infruttuosa: La riforma universitaria: Fuori i mercanti dal tempio... - http://www.cartesio-episteme.net/fuori.html), non resisto alla tentazione di citare l'espressione "analfabeti loquaci" (ma quanto loquaci, in interminabili ed inutili lavori di tante "commissioni" presiedute da altrettanti "presidenti", che con tale ridicolo appellativo soddisfacevano la propria vanità, *todos caballeros*), che è contenuta nel titolo di un ottimo articolo di Roberto Pecchioli (18/12/2022): https://www.ariannaeditrice.it/articoli/gli-analfabeti-loquaci.

Cap. 1 - Eventi, situazioni, probabilità

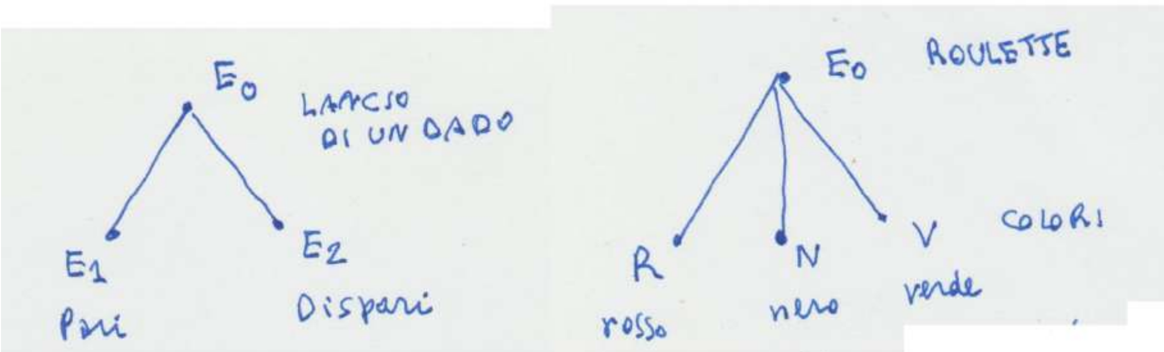
Nel corso di nostre lunghe (ancorché recenti) riflessioni sui fondamenti del calcolo delle probabilità, abbiamo introdotto il termine di **situazione** per definire la coppia formata da un **evento** (concetto che si può ritenere quasi **primitivo**, ma di cui abbiamo comunque tentato una definizione linguistica introducendo il concetto extra-matematico di "isomorfismo semantico") + una sua completa **descrizione**, con il che intendiamo - per cominciare - la precisa elencazione delle **modalità** con cui si può verificare l'evento (le modalità che di esso ci interessano). Le rappresenteremo con **grafi ad albero** di **altezza** 1 (utilizzeremo sempre il simbolo Γ per designare gli alberi che introdurremo, con riferimento alla lettera iniziale di grafo piuttosto che a quella di albero; ovviamente tutti alberi **finiti**, composti cioè da un numero finito di **nodi** = **vertici** e quindi di **rami = lati** - nelle definizioni faremo prevalentemente ricorso ad una metafora botanica auto-esplicativa), ossia, se con E_0 indichiamo l'evento e con $E_1,E_2,,...,E_n$ le dette modalità (va da sé, modalità mutuamente escludentisi) otteniamo il seguente disegnino (che rappresenta quella che diremo una **n -tomia**):



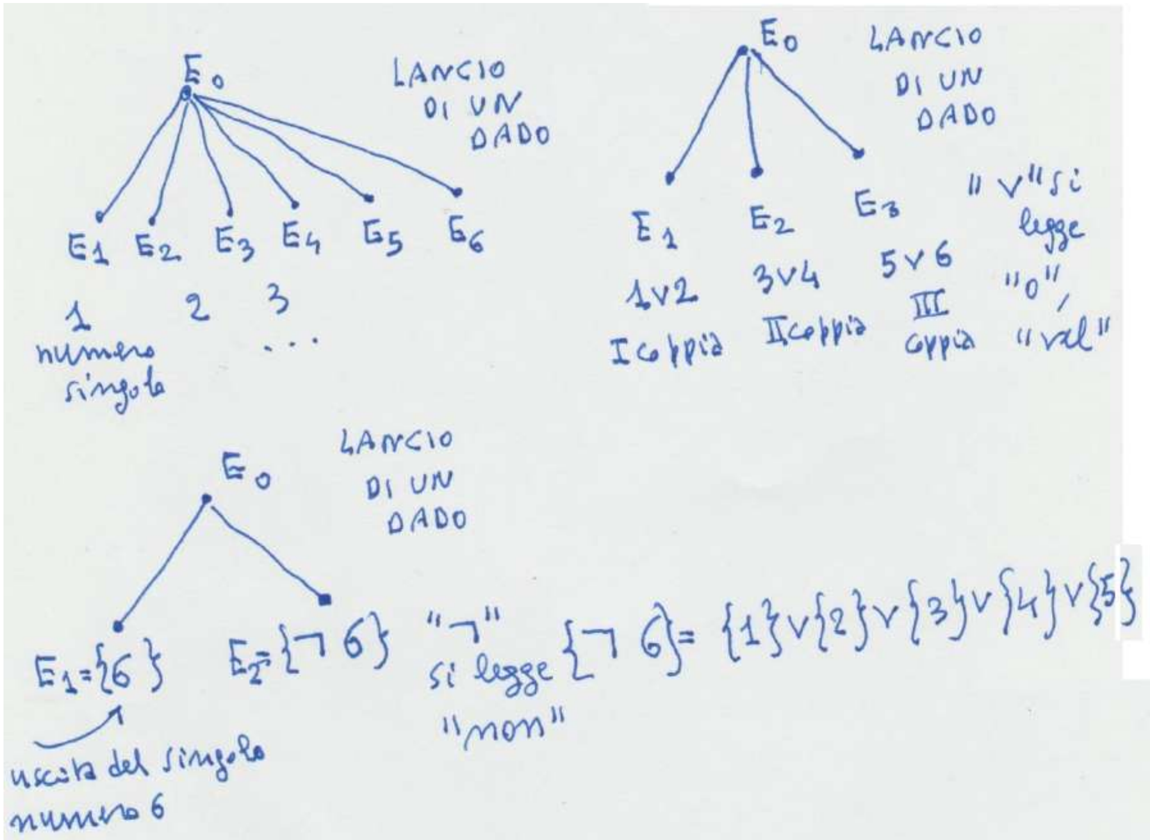
[1]

Il numero naturale n potrà dirsi la **complessità** della situazione in esame.

Esempi banali (evento + descrizione):



[2]



[3]

La coppia evento ed albero **non** basta però per illustrare completamente la situazione, ci vuole anche l'assegnazione di una **funzione di probabilità** definita sui nodi dell'albero, con l'usuale simbolismo matematico:

$$P : \Gamma \rightarrow I_Q$$
$$P(E_0) = x_0\ ,\ P(E_i) = x_i\ ,\ i=1,...,n$$

dove I_Q rappresenta l'intervallo dei numeri razionali (d'onde l'iniziale Q, che rimanda a **quoziente**) compresi tra 0 e 1. Per poter chiamare una tale funzione P una funzione di probabilità (o distribuzione di probabilità) su Γ, introduciamo la sola seguente condizione **C**:

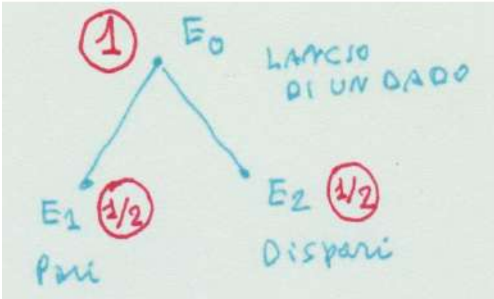
$$\mathbf{C} \qquad P(E_0) = x_0 = 1 = \Sigma(P(E_i)) = \Sigma(x_i),\ i=1,...,n$$
(Σ simbolo di somma)

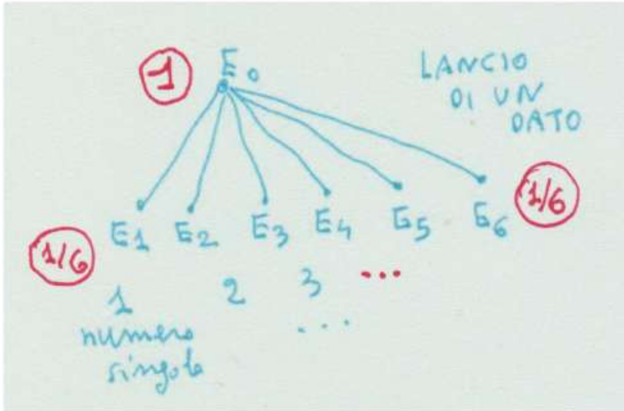
La coppia (Γ,P) diventerà così quello che chiameremo un grafo (albero) **probabilistico**.

La più semplice distribuzione di probabilità è ovviamente:

$\Pi(E_i) = 1/n$ per $\forall$ (ogni) $i=1,...,n$ caso **equiprobabile**

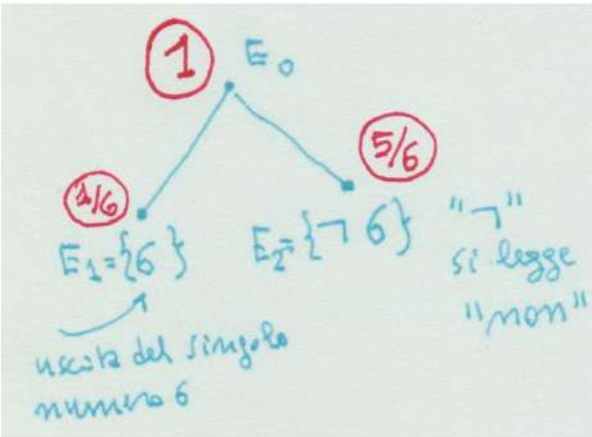
(n, ossia il numero dei rami uscenti dalla **radice** dell'albero, che abbiamo già chiamato complessità dell'albero, può dirsi la **molteplicità** di E_0, mentre E_1 etc. hanno molteplicità 0).
Parleremo della coppia (Γ,Π) come di una n-tomia appunto **semplice** (che rappresenta l'incertezza assoluta), eccone alcuni esempi banali:

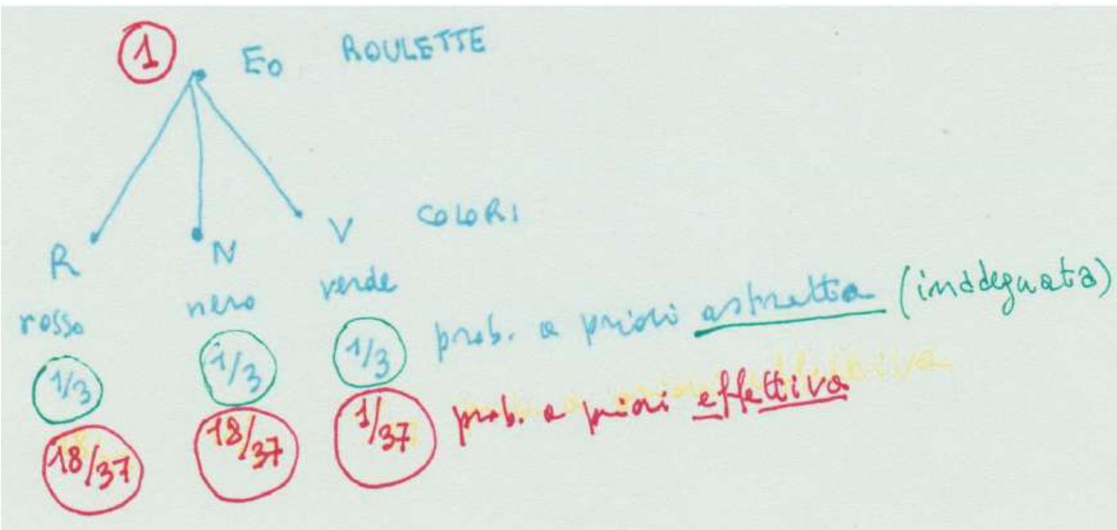




[4]

Potremo dire che una tale distribuzione costituisce la **probabilità a priori astratta** (o **aritmetica**) associata all'albero (se si vuole, $\Pi = P_a$), alla quale sarà sin dall'inizio contrapposta una **probabilità a priori effettiva**, P_e, quella cioè che dipende concretamente dalla situazione in esame, esempi banali:

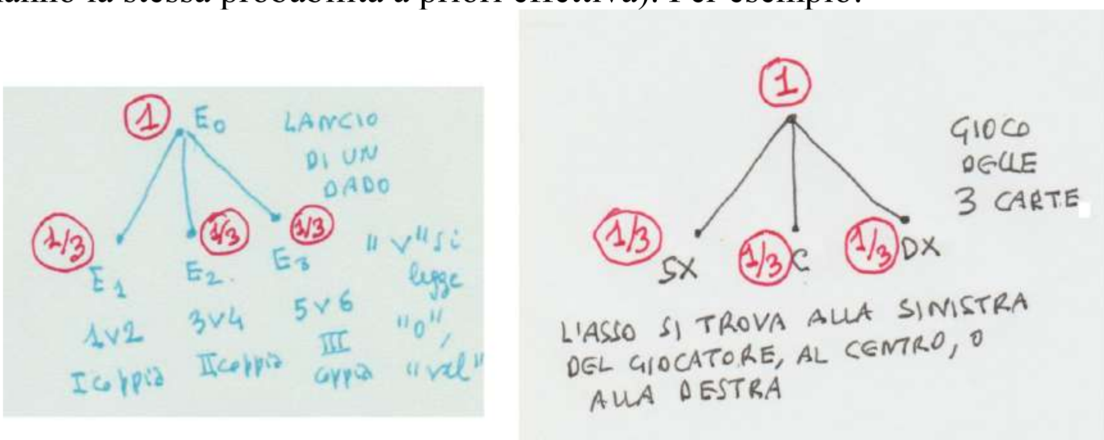




[5]

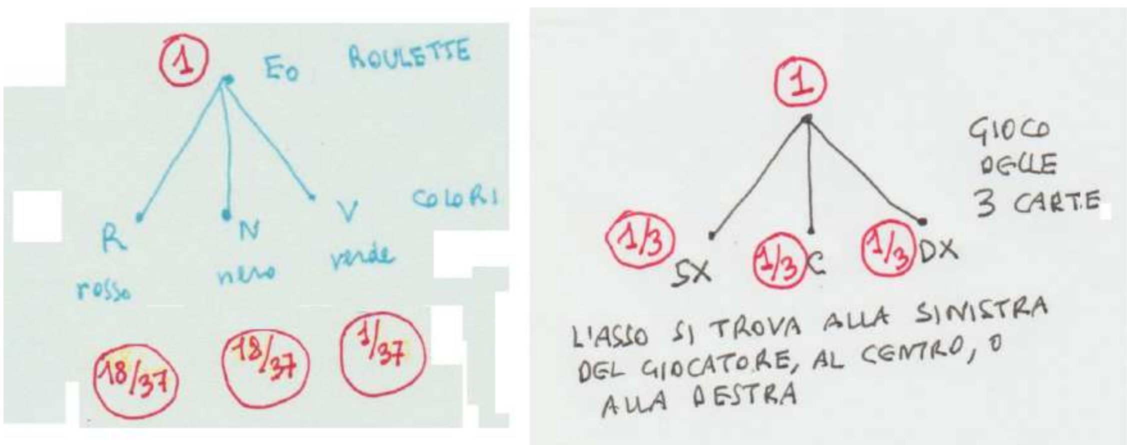
P_e, è una funzione di dominio comune, associata alla situazione sin dall'inizio, e quindi **oggettiva**, anche se, in qualche caso almeno, potrebbe non essere facile riuscire ad individuarla, d'onde la possibilità di errori ed una conseguente presunta condizione di soggettività che invece **non** esiste (vedremo tale possibilità verificarsi proprio nella discussione del paradosso di Monty Hall). Successivamente potranno poi apparire anche **probabilità a posteriori soggettive** (simbolo generico P), dipendenti cioè dallo **stato dell'informazione** dell'**osservatore** in un dato momento t. Nell'istante iniziale t_0 si assume che l'informazione sia soltanto quella (comune) che descrive precisamente la situazione.

Quanto precede consente subito la seguente definizione di **equivalenza** tra situazioni. Due situazioni si dicono equivalenti se hanno la stessa complessità e se i relativi grafi probabilistici (Γ, P_e) sono tra loro **isomorfi** (nell'ovvio senso che, oltre ad essere isomorfi come grafi, risulta pure che nodi corrispondenti hanno la stessa probabilità a priori effettiva). Per esempio:



[6]

sono situazioni ben diverse ma equivalenti, mentre le seguenti, pur essendo descritte da alberi isomorfi, non lo sono:



[7]

Nota 1 (lunga, avanzata) - Nelle pagine da noi dedicate ai fondamenti del calcolo delle probabilità, abbiamo introdotto la quantità

$\Delta(P,P') = \Delta(P',P) = \frac{1}{2} \sum [x_i - x'_i]^2 \quad i=1,...,n$

come misura della **differenza** tra due funzioni di probabilità P e P' definite sullo stesso grafo. Vale a dire, il quadrato normalizzato della **distanza pitagorica** tra il **punto** di coordinate $(x_1,...,x_n)$ che rappresenta P nell'iperpazio Q^n ed il punto $(x'_1,...,x'_n)$ che rappresenta P' nello stesso iperpazio. Δ è un numero razionale compreso tra 0 e 1 (dal momento che la distanza pitagorica massima tra due siffatte funzioni di probabilità è uguale a $\sqrt{2}$, d'onde quel fattore 1/2, e non 1/n come si fa nella definizione classica di **varianza** in statistica), uguale a 0 ses (se e solo se) P=P', mentre la scelta del coefficiente di normalizzazione vale a far sì che Δ sia uguale a 1 (cioè il massimo possibile, ovvero il 100%) ses P e P' sono delle **certezze assolute**, ossia punti del tipo (1,0,0,...,0), (0,1,0,...,0), *etc.*, Tanto per dire, $\Delta(P_e,\Pi)$, dove Π=(1/n,1/n,...,1/n), sarà allora per esempio una misura della **non-equiprobabilità** della n-tomia probabilistica in esame. Un ragionamento analogo conduce alla definizione di **quantità d'informazione** di una distribuzione di probabilità a posteriori P, con la quale si vuole fornire invece una misura dello scarto tra P e il punto P' che possiamo pensare rappresenti l'**incertezza assoluta** della situazione in esame (nel caso astratto sarà il punto dianzi indicato con Π). Più precisamente, potremo definire la **quantità d'informazione astratta** di una distribuzione di probabilità a posteriori P, stenograficamente $QI_a(P)$, come la quantità $QI_a(P) = \sum QI_a(E_i)$, i=1,...,n, dove $QI_a(E_i)$, ossia la **quantità d'informazione astratta** di un singolo nodo E_i della n-tomia in oggetto, sarà definito da:

$$QI_a(E_i) = \frac{n}{n-1}[P(E_i) - \frac{1}{n}]^2 = \frac{n}{n-1}[x_i - \frac{1}{n}]^2 .$$

Qui la scelta del coefficiente di normalizzazione n/(n-1) vale a far sì che $QI_a(P)$ sia sempre minore o uguale ad 1, uguale a 0 ses P coincide con l'incertezza assoluta, ed uguale a 1 (cioè il 100%) ses P è invece una delle **certezze assolute**, Analogamente, potremo parlare della **quantità d'informazione effettiva** di una distribuzione di probabilità a posteriori P, stenograficamente $QI_e(P)$, come della quantità:

$QI_e(P) = \sum QI_e(E_i)$, i=1,...,n, dove adesso invece:

$$QI_e(E_i) = \frac{1}{K}[P(E_i) - z_i]^2 = \frac{1}{K}[x_i - z_i]^2 ,$$

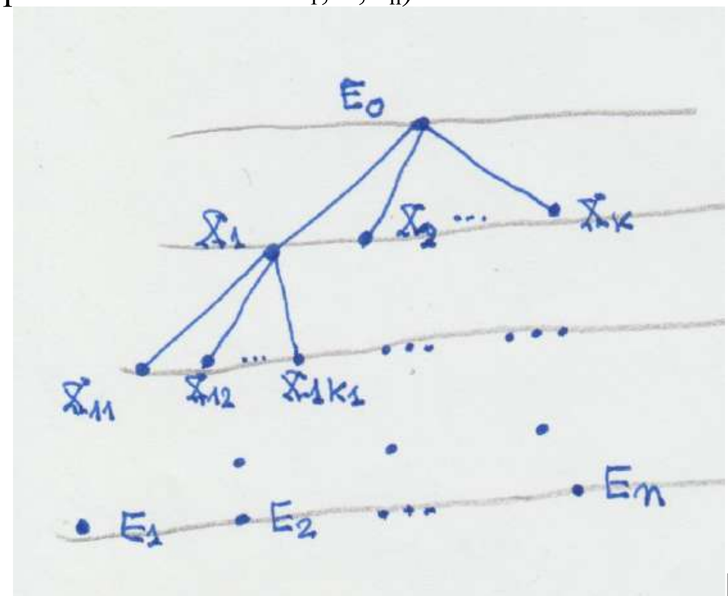
nella quale identità abbiamo indicato con $(z_1,...,z_n)$ le coordinate del punto che rappresenta la corrispondente probabilità a priori effettiva associata all'evento, e con K il quadrato della distanza pitagorica massima tra questo punto $(z_1,...,z_n)$ e le n certezze assolute (nel caso astratto, abbiamo a che fare con n distanze tutte eguali tra loro, e risulta evidentemente K = (n-1)/n).

Concludiamo osservando che, detto P' il punto che rappresenta la detta incertezza assoluta effettiva, sussiste l'identità:

$2\Delta(P,P') = KQI_e(P)$.

Cap. 2 - Decomporre e ricomporre, probabilità relative e probabilità totali

Abbiamo detto che il calcolo delle probabilità a priori effettive è sicuramente oggettivo, ma in qualche caso può non essere immediato, e proprio in tale circostanza consiste come vedremo la difficoltà di ben interpretare il paradosso di Monty Hall. Vediamo però subito come si possa procedere aumentando l'altezza h dell'albero, ossia **diluendo** la descrizione della situazione in diversi passi. In parole povere, cercando di passare da E_0 ad $E_1,E_2,,...,E_n$ attraverso una serie di modalità **intermedie** (sotto-modalità) che possano facilitare il calcolo della distribuzione di probabilità che ci interessa (potremmo anche dire di aver trasformato la complessità **orizzontale** della situazione in esame in una complessità verticale, più facile però da gestire). L'albero rappresentato nel seguente disegnino dovrebbe chiarire il procedimento descritto (il **livello** 0 contiene solo E_0, il livello 1 o primo livello contiene $X_1,...,X_k$, *etc.*, fino al livello h-mo in cui appaiono finalmente $E_1,...,E_n$):

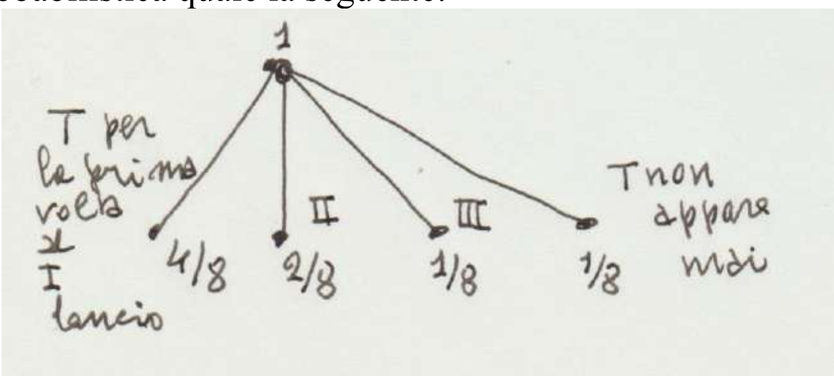


[8]

Nota 2 - Gli alberi con i quali avremo a che fare d'ora in avanti hanno una struttura naturale che si dice di **ordine parziale**, nel senso che, per due nodi μ e ν dell'albero, avrà significato scrivere μ ≤ ν oppure μ ≥ ν, a seconda che μ **preceda** o **segua** ν. Una **relazione** di sicuro **anti-simmetrica** (se μ ≤ ν e μ ≥ ν allora necessariamente μ = ν), **riflessiva** (μ ≤ μ) e soprattutto **transitiva** (μ ≤ μ' e μ' ≤ μ'' implica μ ≤ μ''), e parziale nel senso che qualche volta non varranno né μ ≤ ν né μ ≥ ν, nel qual caso si dirà che i due nodi μ, ν sono tra loro **inconfrontabili**. Detto che questi alberi descrivono in qualche modo il lavoro dell'intelletto di uno scacchista, quando comincia ad esaminare certe possibilità per una mossa, e tiene poi conto per ciascuna di esse di una possibile contro-mossa dell'avversario, e così via, chiudiamo la Nota osservando che si tratta di strutture di ordine parziale alquanto particolari, che si chiamano talvolta **semi-reticoli**. Esse godono della proprietà che, comunque assegnati due nodi μ, ν, esiste un unico nodo che sia al tempo stesso ≤ sia di μ che di ν, e sia **massimo** rispetto a queste due proprietà. Insomma, una sorta di **massimo comun divisore** di μ e ν (eventualmente coincidente con la radice dell'albero), che si può indicare con il simbolo insiemistico d'intersezione, ossia μ∩ν. Il termine **reticolo** viene riservato a quelle strutture d'ordine parziale che ammettono anche un'**operazione** di **unione**, la quale aritmeticamente corrisponderebbe ad un **minimo comune multiplo**, un nodo che negli alberi però non esiste quasi mai, a meno che non si tratti di una coppia di nodi tra loro **confrontabili**, appartenenti cioè ad un unico **stelo**.

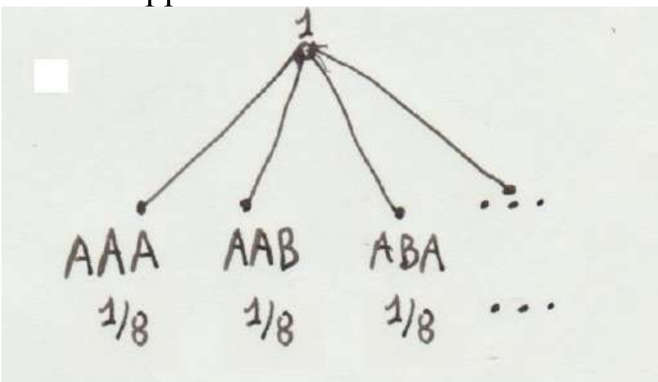
La complessità ultima della situazione rimane ovviamente invariata, anche se potrebbe apparire addirittura aumentata, tenendo conto che avremo più nodi da prendere in considerazione, nodi però singolarmente più "facili" da trattare sotto il profilo probabilistico degli ultimi, quelli **massimali**. In altre parole, il miglioramento consiste nella maggiore facilità di individuare le probabilità di **tutti** i singoli nodi, ivi compresi quelli che ci interessavano all'inizio, cioè gli ultimi, quelli che nell'albero si trovano **più in alto** - o meglio **più in basso** secondo la convenzione grafica da noi prescelta, ordine dall'alto verso il basso del foglio, radice in alto. Ma. ripetiamo, una cosa è la complessità dell'albero, ed un'altra quella della situazione in esame. Precisamente, se h è l'**altezza** dell'albero, potremo parlare della **complessità dell'albero** come della (h+1)-pla ordinata $(n_0=1,n_1,n_2,...,n_h)$: ovviamente, $n_0 \leq n_1 \leq n_2 \leq ... \leq n_h$, $n_h = n$, numero degli elementi massimali dell'albero, cioè quel numero che abbiamo chiamato **complessità della situazione**. Manifestamente, $1+n_1+n_2+...+n_h$ è il numero totale dei nodi dell'albero, mentre $n_1+n_2+...+n_h$ è il numero totale dei rami dell'albero.

Cerchiamo di chiarire tutto ciò che precede con un facile esempio. Supponiamo di lanciare una moneta per 3 volte di seguito, e di chiedere poi quale sia la probabilità (a priori effettiva!) che la testa esca per la prima volta al primo lancio, per la prima volta al secondo lancio, etc., fino a quella di non uscire mai nel corso dei 3 lanci (complessità della situazione uguale a 4). Non si tratta di un calcolo difficile, per esempio si possono prima di tutto determinare tutte le possibili "schedine" che descrivono l'esito dei 3 lanci, TTT, TTC, *etc.* (T=testa, C=croce), schedine che si trovano essere in numero di 8 ($8=2^3$), e raggruppare poi una ad una quelle che presentano T per la prima volta al primo posto, quelle che presentano T per la prima volta al secondo, e così via. Otterremo allora una 4-tomia probabilistica quale la seguente:



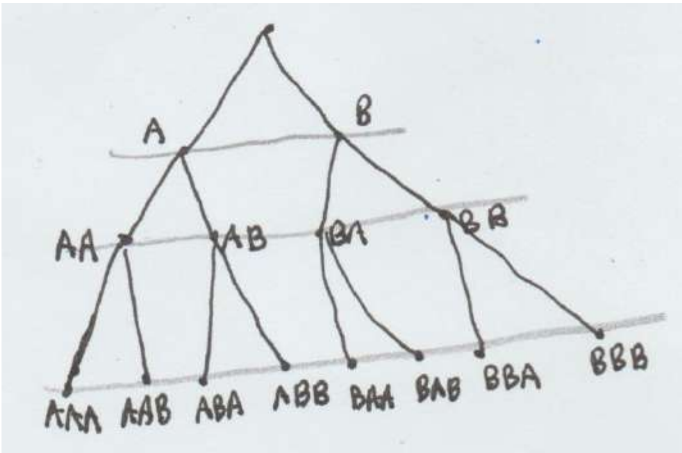
[9]

La correttezza del calcolo è verificata dall'identità $\frac{4}{8}+\frac{2}{8}+\frac{1}{8}+\frac{1}{8}=1$, e fin qui tutto bene. Ma abbiamo introdotto un "automatismo" che ci consente di arrivare al risultato senza nessuna fatica intellettuale, e soprattutto senza commettere errori nel corso del cammino? E' prima di tutto nostra lunga esperienza didattica che molte persone omettono qualcuna delle dette 8 terne TTT, *etc.*, quando le vanno ad elencare, per cui è preferibile prima di tutto usare al posto dei simboli T,C i simboli A,B rispettivamente, in modo da potersi basare sull'**ordine alfabetico**. La 8-tomia che descrive i risultati dei 3 lanci consecutivi (ossia le dette schedine), si può allora rappresentare come:



[10]

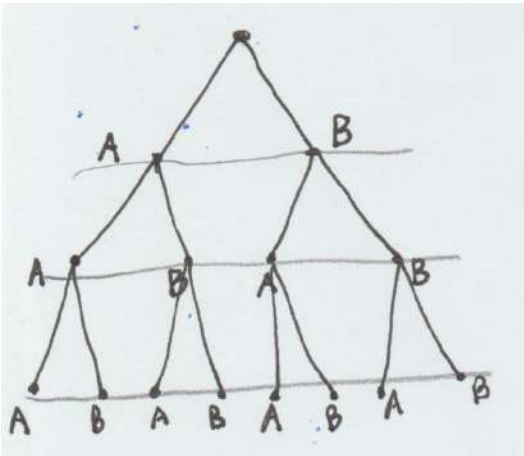
ovvero, **diluendo**:



[11]

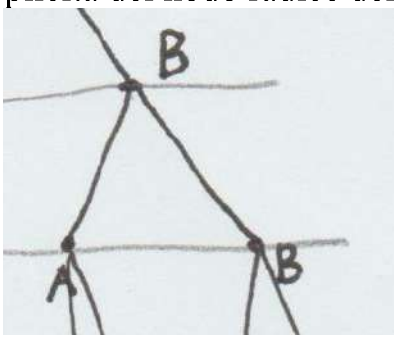
(un grafo di altezza 3 che potremo dire **uniforme** di **caratteristica** 2, nel senso che la molteplicità di ogni nodo, esclusi quelli massimali, è uguale appunto a 2; complessità dell'albero = $(1,2,2^2,2^3)=(1,2,4,8)$).
Le 4 uscite di A per la prima volta al I posto sono i primi 4 elementi massimali del precedente grafo, leggendo l'ultima riga da sinistra a destra, le 2 uscite di A per la prima volta ai II posto di trovano subito dopo, l'unica con A per la prima volta al III posto come penultima, ed infine nessuna uscita di A all'ultimo posto.

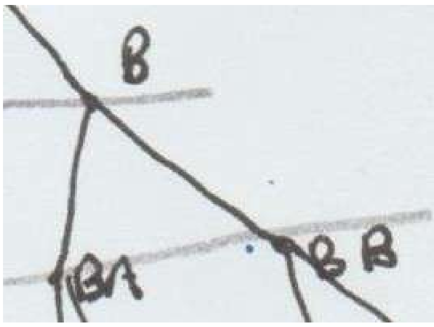
Prima di procedere oltre, sottolineiamo come si possa addirittura semplificare la diluizione indicando ad ogni **passo** (nodo) soltanto le due possibilità A,B che dopo di esso si aprono, ossia scrivendo:



[12]

Pure questo è un albero dal significato del tutto auto-evidente, dove appaiono invero tanti nodi **diversi** indicati però con lo **stesso** simbolo, una scelta la quale farà sì che, quando andremo a parlare di una distribuzione di probabilità ad esso associata, diciamola P, se volessimo per esempio parlare di una certa P(A), non sapremmo di quale A si tratta: l'unico A che si trova al I livello, oppure tanto per dire quello che si trova al II livello nella terza **sezione** dell'albero? (contando dall'alto verso il basso secondo il disegno, e da sinistra a destra; con il termine sezione intendiamo il sotto-albero costituito da un nodo non massimale e da **tutti** i nodi ad esso **successivi**, ovviamente ogni sezione è una m-tomia, dove m è la molteplicità del nodo radice della sezione).

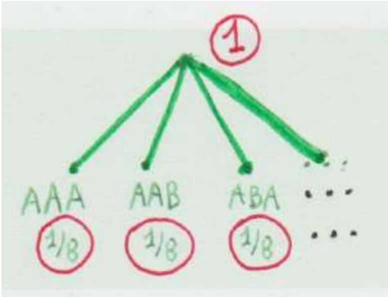


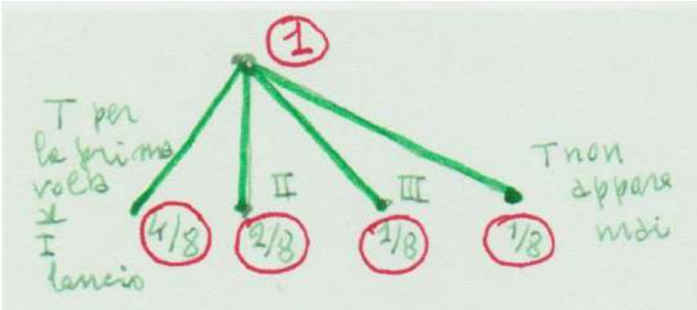


[13]

3a sezione semplificata 3a sezione non semplificata

Quello che precede è un problema di notazione che non era stato ben spiegato nella precedente versione di questo scritto, e che si potrebbe però sempre risolvere in maniera suggestiva inserendo i valori di probabilità all'interno di circoletti contigui ad ogni vertice dell'albero, per esempio colorandoli di rosso (lo abbiamo già fatto nel capitolo precedente, una pratica comunque ... dispendiosa, che quindi non seguiremo sempre, introducendo così ulteriori semplificazioni, almeno quando esse impediscano un'interpretazione univoca):

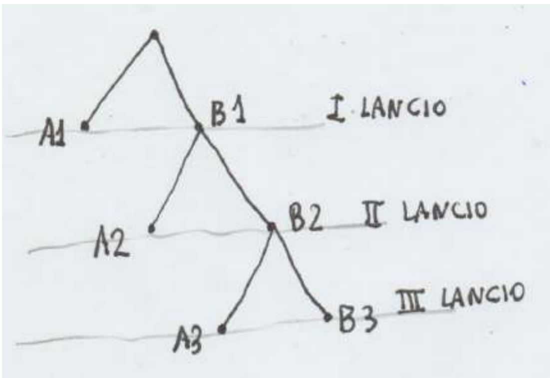




[14]

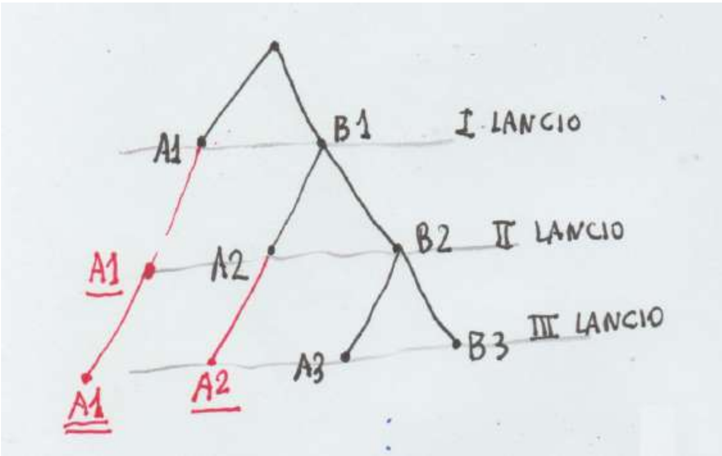
Ecco che viene così alla mente l'immagine dell'**albero di Natale** richiamata nel titolo, rafforzata volendo da un eventuale utilizzo del colore verde nel disegno dell'albero - disegni che in effetti diventano ancora più suggestivi nel caso di alberi "alti". Notiamo però che, se è vero che la somiglianza con gli alberi appare più evidente se "rovesciamo" la convenzione di riportare la radice nella parte superiore dei nostri grafici, è anche vero che, parlando di "alberi di Natale", conviene invece mantenerla, sebbene la radice in questo caso assomigli invero più alla "punta" dell'albero che non alla sua base.

Tornando al problema in esame, notiamo come si possa addivenire alla conclusione ancora più rapidamente introducendo il seguente albero **a cascata di dicotomie**:



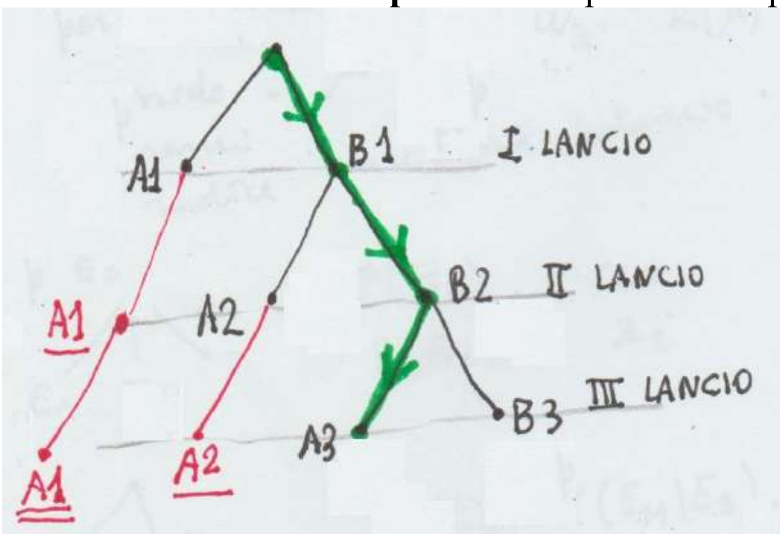
[15]

un grafo di altezza 3 non regolare, dove con l'appellativo **regolare** indichiamo la circostanza che tutti gli elementi massimali dell'albero sono alla stessa "distanza" dalla radice. Un grafo non regolare si può evidentemente sempre **regolarizzare** nel modo illustrato dal seguente disegno, aggiungendo cioè all'albero dei semplici **steli**, vale a dire attribuendo molteplicità 1 a tutti quegli elementi massimali che hanno distanza dalla radice minore dell'altezza dell'albero:



[16]

Otteniamo così subito all'ultimo livello dell'albero le 4 modalità della situazione che ci interessavano (A1 rappresenta l'uscita di A per la prima volta al primo lancio, A2 l'uscita di A per la prima volta al II lancio, A3 l'uscita di A per la prima volta al III lancio, B3 la schedina BBB, cioè nessuna uscita di A. Basta seguire i **percorsi**, o **cammini orientati**, dalla radice ai vari elementi massimali dell'albero, ossia effettuare una **ricomposizione** dopo la decomposizione;



[17]

(detta per esempio R la radice dell'albero, il cammino in oggetto è R-B1-B2-A3, ossia l'**unico** cammino che va da R ad A3)

All'albero sopra riportato per descrivere la situazione in parola, mancano però ancora le probabilità, sicché vogliamo affrontare a termine di capitolo il problema della definizione di una distribuzione di probabilità per alberi di **altezza superiore ad 1**. Nel farlo, ci troveremo di fronte ad una **doppia possibilità** che vogliamo ben chiarire alla nostra mente.

Introduciamo prima di tutto la banale nozione di **distanza** di un nodo μ di Γ (scegliamo il simbolo μ con riferimento alla lettera iniziale di "modalità"), distanza ovviamente relativa a Γ, come la lunghezza dell'unico cammino orientato che dalla radice di Γ va fino al nodo in questione, e torniamo poi al concetto di sezione S di Γ. Detta distanza di S (in Γ) la distanza della radice di S, è evidente che due sezioni equidistanti sono tra loro disgiunte, e che altrettanto disgiunte sono le sezioni la differenza delle cui distanze è ≥ 2. L'unico caso in cui l'intersezione di due sezioni S, S' non sia vuota, è quello in cui trattisi di due sezioni **consecutive**, e tali che, supponendo per esempio, con notazione autoevidente S < S', la radice di S' sia uno dei nodi di S diverso ovviamente dalla radice di S (due sezioni distinte non possono mai avere la radice comune). Insomma, un albero di altezza superiore ad 1 si può pensare come il risultato di un opportuno molteplice **incollamento** tra loro di tante diverse k-tomie, per ciascuna delle quali abbiamo già un concetto di probabilità, ossia una qualsiasi funzione (del tipo precisato nel Cap. 1) che soddisfa la condizione **C**. Indipendentemente assegnate tante funzioni di probabilità per ciascuna sezione, otterremo tante k-tomie probabilistiche, che si tratta adesso di incollare fino ad ottenere un'unica funzione di probabilità definita sull'intero Γ. Tornando al caso delle due dicotomie di prima, S ed S', è chiaro infatti che il nodo μ di S che è radice di S' avrà probabilità 1 in quanto radice di S', e probabilità quella che abbiamo deciso in quanto nodo di S. Bene, incolliamo le sezioni lasciando da parte tutti gli 1 che vengono dai nodi di sezioni che diventano radici di una sezione successiva, ed avremo ottenuto quella che potremo dire una **distribuzione di probabilità relative** P_ρ definita sull'intero su Γ (e potrebbe poi trattarsi di probabilità a priori oggettive, astratte o effettive, oppure a posteriori soggettive, dipendenti cioè dallo stato dell'informazione dell'osservatore che le introduce, non c'e nessuna differenza per ciò che concerne le loro proprietà formali).

In altre parole, si può immaginare che, percorrendo un certo cammino orientato, ci si trovi in un certo momento fermi su un nodo, davanti al quale (se non è un nodo terminale) si aprono diverse possibilità di andare avanti, esattamente m se m è la molteplicità del nodo, e quindi che, ad ogni possibilità di procedere oltre (ovvero, per ciascuno dei rami che partono dal nodo procedendo verso il basso) corrisponda una probabilità che abbiamo appunto detta relativa.

Chiaramente, avremo a che fare come in precedenza con una funzione

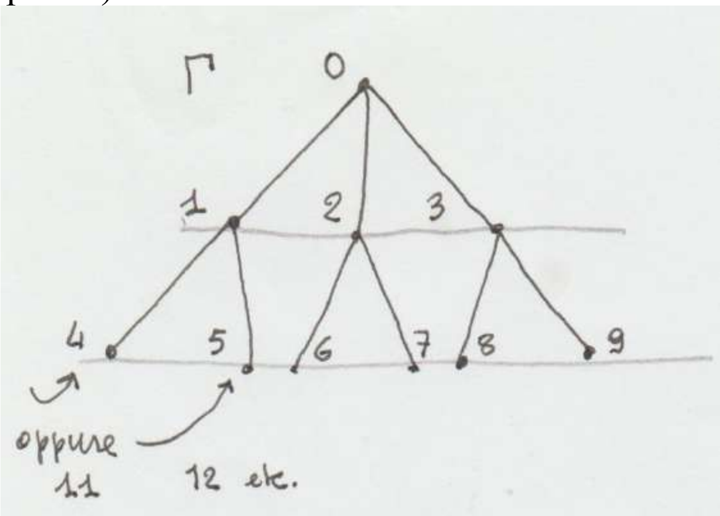
$$P_\rho : \Gamma \rightarrow I_Q$$

la quale soddisfa, al posto della condizione **C** di cui al cap. 1, la seguente condizione **R**:

$$\mathbf{R} \qquad P_\rho(\text{radice di } \Gamma) = 1 = \Sigma(P_\rho(E_i)),\ i=1,...,k$$

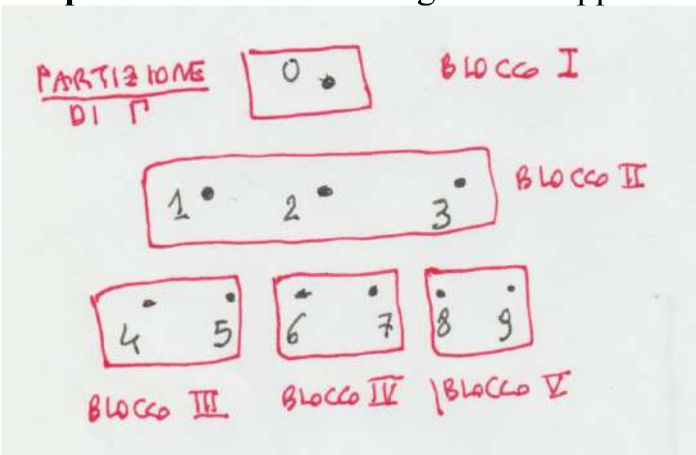
dove con $E_1,E_2,,...,E_k$ abbiamo indicato una qualsiasi k-tomia sezione dell'albero Γ. Nel caso di un albero di altezza 1, che sia cioè lui stesso una k-tomia, e quindi pure l'unica sezione dell'albero, la condizione **R** coincide con la condizione **C**.

Volendo chiarire ulteriormente la situazione, eliminiamo i vertici da ogni sezione, pur continuando a prendere in considerazione la radice di Γ come un punto isolato. Passeremo così per esempio da un grafo regolare del tipo seguente (un grafo che potremo dire **localmente uniforme**, perché non è globalmente uniforme, ma almeno tutti i nodi che hanno la stessa distanza dalla radice hanno la stessa molteplicità):



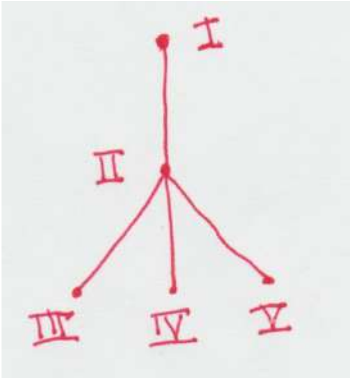
[18]

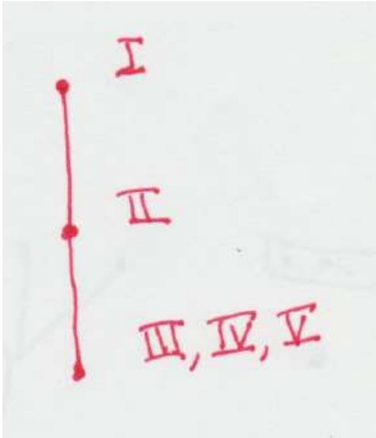
ad una ben precisa **partizione** di Γ il cui significato appare del tutto manifesto:



[19]

Nota 3 - Una partizione che anch'essa può andare a costituire un albero, naturalmente assai **ridotto** rispetto al precedente (una sorta di **scheletro** dell'albero di partenza), come nel naturale disegno qui di seguito riportato a sinistra (a destra invece l'albero che si ottiene a sua volta dallo scheletro di Γ iterando il procedimento):

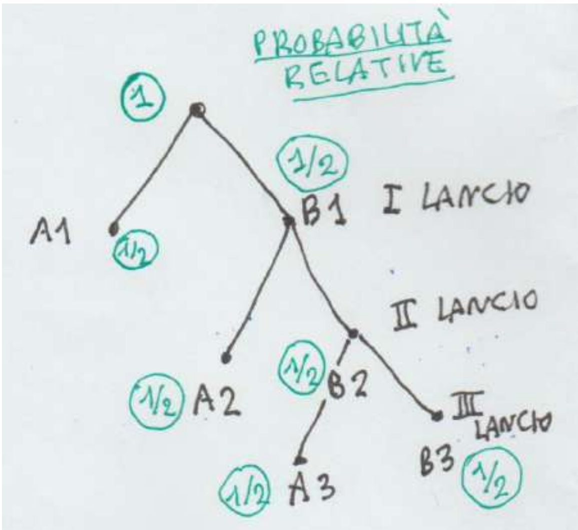




[20]

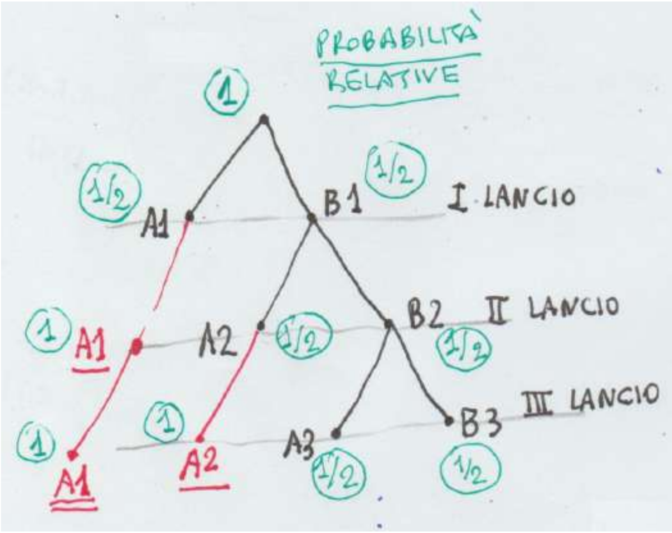
Nota 4 - Val forse la pena di ricordare che una **partizione** di un dato insieme (non vuoto) A è un insieme di sottoinsiemi di A a due a due disgiunti, e tali che l'unione di essi costituisca l'intero A (si parla di k-partizione se si ha che fare con k sottoinsiemi). Si aggiunge di solito la condizione che tra gli elementi della partizione non ci sia l'insieme vuoto, ma noi utilizzeremo il termine partizione in senso lato, ossia contempleremo eventualmente con il termine partizione pure questo caso (per noi corrispondente ad una probabilità uguale a 0 - si veda il capitolo successivo).

Questa che segue è per esempio la **distribuzione di probabilità relative a priori effettive** P_ρ definita sul grafo Γ a cascata di dicotomie dianzi introdotto per descrivere una situazione inerente a 3 lanci consecutivi di 3 monete;



[21]

mentre questa è la "stessa" distribuzione sull'associato grafo "regolarizzato":



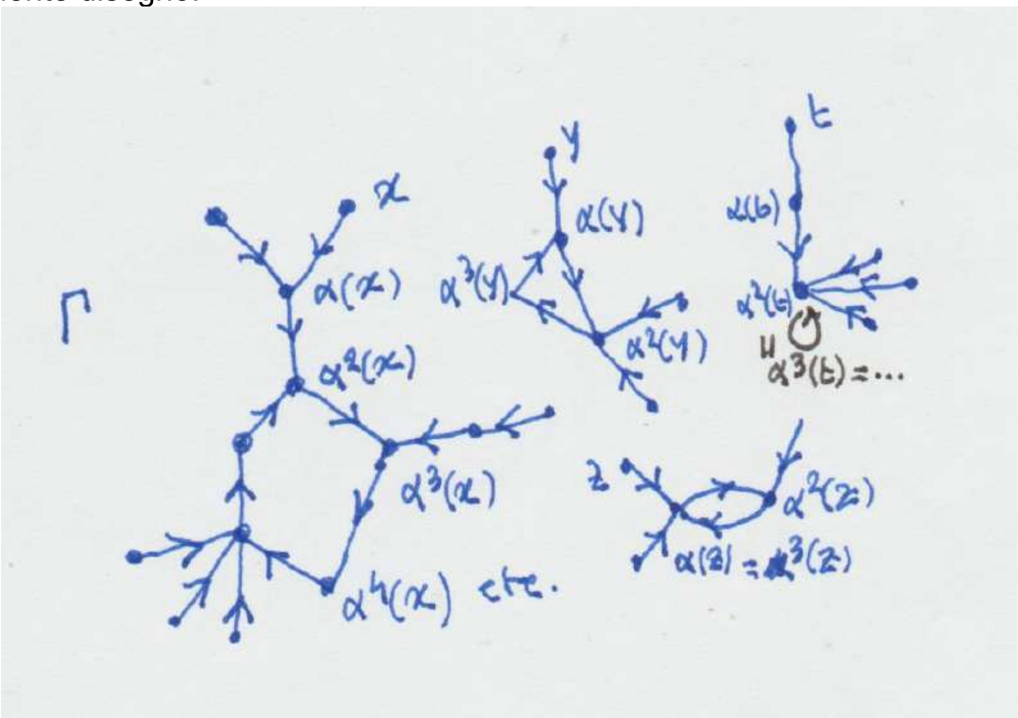
[22]

Abbiamo inserito adesso le probabilità all'interno di circoletti di colore verde, scegliendo così un modo per distinguerle dalle corrispondenti probabilità **totali** (in altre nostre note le abbiamo dette **assolute**, ma preferiamo adesso mantenere la scelta di riservare l'iniziale "a" all'aggettivo "astratta"), per le quali utilizzeremo invece (se ne vedrà un esempio tra poco) circoletti di colore rosso. Per questo secondo tipo di probabilità introdurremo il simbolo P_τ, ed a P_τ vogliamo infine arrivare a partire da una data P_ρ. Val la pena di ripetere, o di ulteriormente sottolineare, che stiamo procedendo in questa maniera che può apparire eccessivamente laboriosa perché, se è vero che ci interessa principalmente P_τ - soprattutto all'ultimo livello, non dimentichiamolo - è altrettanto vero che l'individuazione di P_ρ è in molti casi del tutto semplice e diretta (automatismo), così come poi sarà altrettanto semplice diretto il passaggio conclusivo da P_ρ a P_τ (vale a dire, un primo automatismo seguito poi da un secondo, ce ne renderemo ben conto quando nell'ultimo capitolo discuteremo finalmente il paradosso di Monty Hall).

Ricapitolando, eccoci giunti alla questione teorica fondamentale di questo capitolo: **come si passa da da P_ρ a P_τ?**

La risposta è fortunatamente facile, e ci sono diversi modi per giustificarla. Basterebbe forse notare che, se $P_\rho(\mu)$ - per esempio 1/2 - è la probabilità di un nodo μ diverso dalla radice di Γ, si tratta di un'informazione riguardante μ ma relativa al nodo **antecedente** di μ (un nodo che potremo suggestivamente indicare con il simbolo μ-1), ossia alla radice della **prima** sezione di Γ di cui μ fa parte (procedendo al solito dal **basso** dell'albero verso l'**alto**, basso ed alto che appaiono però sempre rovesciati secondo la convenzione prescelta per le nostre rappresentazioni grafiche). Ogni nodo diverso dalla radice ammette infatti un **unico antecedente**, e volendo m **successivi**, se m è la molteplicità del nodo (eventualmente m=0, ciò che accade ses μ è un elemento massimale di Γ).

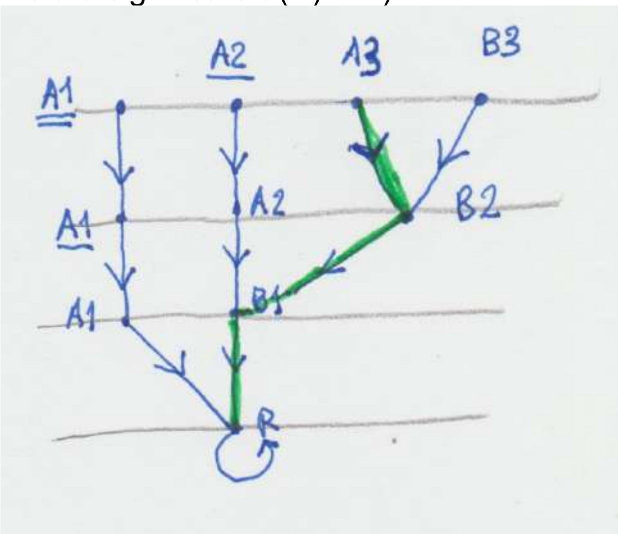
Nota 5 (avanzata) - La proprietà dell'albero che abbiamo appena evidenziato, ossia quella relativa all'esistenza di un unico antecedente, permette di introdurre una connessione tra **teoria degli alberi** e **teoria dei sistemi dinamici** (o **semi-dinamici**, come preferiamo dire) **discreti**, che va al di là delle finalità del presente scritto, ma alla quale ci pare comunque istruttivo (e "divertente") accennare. Nel contesto in cui ci troviamo, è possibile infatti definire su Γ (sul "sostegno" di Γ, volendo essere precisi, ossia su Γ come puro insieme e non come "struttura"; si tratta di un termine peraltro già usato in precedenza senza allora però nemmeno un cenno di spiegazione) una particolare funzione α che va da Γ a Γ (simili funzioni sono dette in generale **endomorfismi**, o anche **semi-permutazioni**), $\alpha : \Gamma \rightarrow \Gamma$, che associa ad ogni nodo diverso dalla radice il proprio antecedente, ed alla radice se stessa. **Ogni** tale endomorfismo induce su Γ una "struttura" di grafo del tipo di quella raffigurata nel seguente disegno:



[23]

(un endomorfismo α definito su un insieme con 29 elementi, che induce su Γ una 4-partizione non uniforme, ossia 4 **attrattori** di diverse **lunghezze** 5, 3, 1, 2),

struttura che si può interpretare suggestivamente così. Un certo elemento x di Γ "**va in**" α(x), α(x) va a sua volta in $\alpha(\alpha(x)) = \alpha^2(x)$, e così via, come se si trattasse appunto di un "movimento" indotto da α sugli elementi di Γ. Una descrizione alternativa introduce una funzione $\alpha : N'x\Gamma \rightarrow \Gamma$, dove N' è l'insieme dei numeri naturali più lo zero (volendo), e $\alpha(n,x) = \alpha^n(x)$ - ovviamente: $\alpha(m+n,x) = \alpha(m,\alpha(n,x))$. Essendo Γ un insieme finito, è evidente che ad un certo punto la successione $\alpha(x)$, $\alpha^2(x)$, ... deve **tornare su se stessa**, $\alpha^k(x) = \alpha^{k+p}(x)$, e che a partire da quel punto in poi deve cominciare a percorrere dei **cicli** (**attrattori**) dai quali, una volta **finitici dentro**, gli elementi di Γ non si potranno staccare mai più. Si ottiene in tal modo una partizione di Γ, che si dirà un albero ses esiste un unico attrattore di lunghezza 0 (**punto fisso**), una descrizione diversa e molto interessante della medesima fenomenologia matematica (concettuale). Nel seguente disegno un esempio del sistema semi-dinamico discreto corrispondente all'albero della figura [16], evidenziato in verde il cammino **opposto** a quello riportato con il medesimo colore nella figura [17] (basta "rovesciare l'albero e procedere ancora dall'alto verso il basso; notiamo che abbiamo adesso esplicitamente indicato con R la radice dell'albero, e che la freccia circolare apposta vicino ad R sta a significare α(R) = R):



[24]

Chiudiamo la presente lunga Nota accennando alla circostanza che abbiamo chiamato il sistema in oggetto semi-dinamico volendo riservare il termine **dinamico** a funzioni α : ΓxZ → Γ (dove Z è l'insieme dei numeri interi relativi, dall'iniziale del termine tedesco *Zahl*, numero), ossia numeri interi con segno, Funzioni che corrisponderanno manifestamente ad endomorfismi biunivoci di Γ (**permutazioni**). I sistemi semi-dinamici descrivono soltanto movimenti nel **futuro**, quelli dinamici anche nel **passato**, *etc. etc.*, un argomento *a latere* di quello qui trattato che potrebbe essere in ogni caso interessante sviluppare un poco.

Tornando al nostro discorso principale, se μ non ha molteplicità 0, μ farà parte anche di una **seconda** sezione, che potremo dire **successiva** a quella di cui abbiamo parlato, e di una soltanto, sezione di cui μ costituirà adesso la radice, e quindi con probabilità relativa (**locale**) uguale ad 1. Orbene, se μ-1 ha probabilità totale in Γ uguale per esempio a 1/3, allora la corrispondente probabilità totale di μ in Γ dovrà essere intesa come 1/2 di 1/3, la metà di una terza parte, vale a dire 1/6. Ovvero, la probabilità totale $P_\tau(\mu)$ di μ in Γ derivante dalla distribuzione di probabilità relativa P_ρ, può essere definita dalla seguente formula **ricorsiva**:

$P_\tau(\mu) = P_\rho(\mu)P_\tau(\mu\text{-}1)$,

ricorsiva nel senso che $P_\tau(\mu\text{-}1)$ sarà sua volta individuato da $P_\rho(\mu\text{-}1)P_\tau(\mu\text{-}2)$ *etc.*, e quindi la precedente formula diventa:

$P_\tau(\mu) = P_\rho(\mu)P_\rho(\mu\text{-}1)P_\tau(\mu\text{-}2)$,

e così via a ritroso, fino ad arrivare alla radice dell'albero, per quale sia P_ρ che P_τ valgono 1. Otteniamo così ancora una funzione

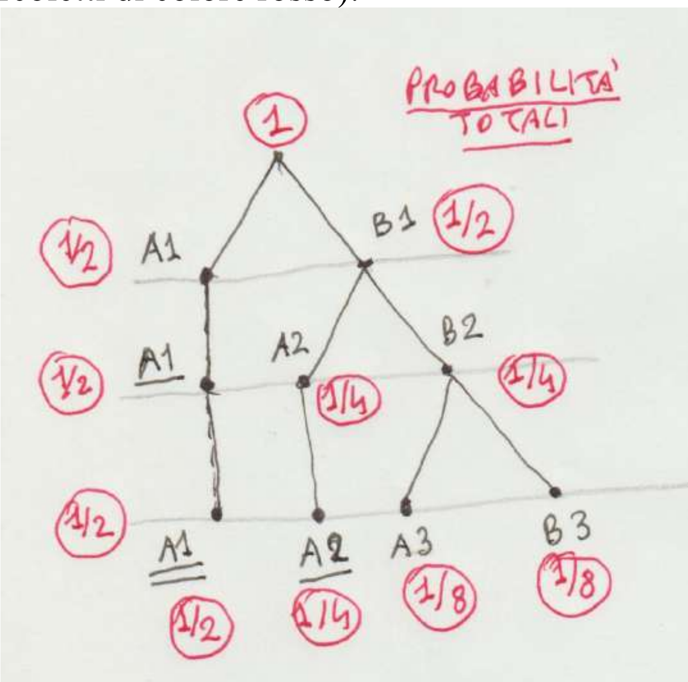
$$P_\tau : \Gamma \rightarrow I_Q$$

la quale però, in luogo della condizione **R** di cui sopra, soddisfa invece alla seguente condizione **T**:

T $P_\tau(\text{radice di } \Gamma) = 1 \; ; \; P_\tau(X_0) = \Sigma(P_\tau(X_i)), \; i=1,\ldots,k$

dove con $X_0, X_1, X_2,, \ldots, X_k$ è stata indicata una qualsiasi k-tomia sezione dell'albero in esame (è chiaro adesso perché si siano utilizzate le due iniziali **R** e **T** per definire le proprietà aritmetiche fondamentali che devono essere soddisfatte da una distribuzione di probabilità, relative o totali, su Γ). Nel caso di un albero di altezza 1, che sia cioè lui stesso una k-tomia, e quindi anche l'unica sezione dell'albero, le due nozioni di probabilità relative e probabilità totali manifestamente coincidono.

Per chiudere un discorso finora rimasto aperto, ecco finalmente qui di seguito il grafo probabilistico (un po' "raddrizzato" rispetto al precedente) che abbiamo introdotto per descrivere una situazione inerente a 3 lanci consecutivi di 3 monete con le sue probabilità totali (inserite stavolta, come già annunciato, all'interno di circoletti di colore rosso):



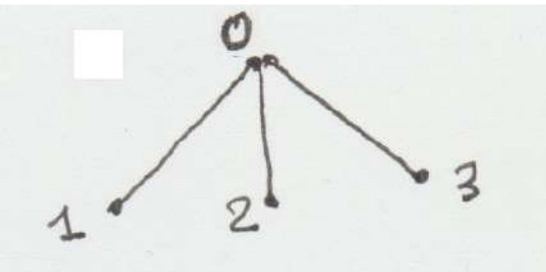
[25]

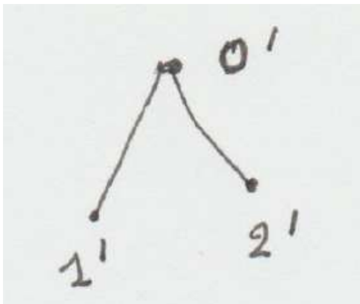
Insomma, lungo un cammino orientato quale per esempio il cammino R-B1-B2-A3 evidenziato in verde nella figura [17] (e, ricordiamo, R = radice dell'albero) le diverse probabilità relative che si incontrano ad ogni nodo del cammino vanno man mano moltiplicate tra loro (potremmo dire suggestivamente che ogni probabilità totale in un dato nodo determinata comprenda in sé tutto il **passato** che l'ha preceduta), fino ad individuare una successione decrescente (ma non sempre **strettamente** tale, se qualcuna delle probabilità relative incontrate è uguale ad 1), nel caso in esame 1 > 1/2 > 1/4 > 1/8.

Nota 6 - Si osservi bene che, mentre il passaggio da P_ρ a P_τ è univocamente determinato, quello inverso non lo è altrettanto, dal momento che un'identità del tipo $P_\rho(\mu) = P_\tau(\mu)/P_\tau(\mu-1)$ non ha significato in presenza di probabilità totali che siano uguali a 0. Aggiungiamo che la questione dell'**emergenza** dei due casi limite 0 o 1 è interessante in sé, e che non appare conveniente escluderli per definizione, e non solo nel caso di probabilità a posteriori soggettive, ma anche di probabilità a priori effettive (un caso in cui di fronte ad uno 0 potrebbe venire la tentazione di cancellare il ramo che ad esso porta), come vedremo anche nella nostra discussione del paradosso di Monty Hall. Qui aggiungiamo solo che le probabilità totali vanno ovviamente a diminuire lungo ogni cammino orientato che dalla radice di un albero conduce verso uno dei suoi nodi massimali, sicchè **dopo** uno 0 non può che esserci 0 (lo 0 per così dire "contagia" tutti i nodi con cui interagisce nel futuro), mentre **prima** di un1 non ci possono essere che 1 (potremo parlare di un **punto fermo**).

Detto che nel prossimo capitolo discuteremo un interessante esempio "universale" della probabilità, un'interpretazione insiemistica che connette immediatamente **probabilità** e **statistica**, e che chiarirà in altro modo il passaggio da P_ρ a P_τ, chiudiamo quello presente con due ulteriori Note, interessanti ma un po' più avanzate rispetto alla maggior parte degli argomenti fin qui discussi.

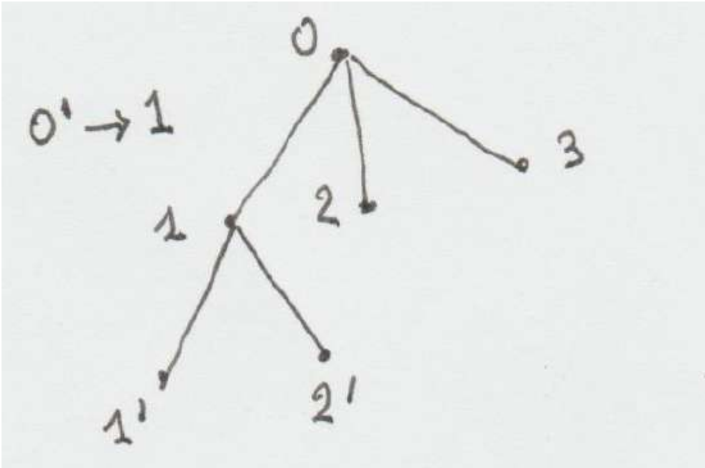
Nota 7 - Vogliamo approfondire adesso il concetto di "incollamento" di una k'-tomia in una k-tomia quando entrambe siano però x-tomie **probabilistiche**, una questione che illustrerà una volta di più il significato della regola di passaggio da P_ρ a P_τ, (un'altra di natura insiemistica la vedremo nel capitolo successivo). Si parta per esempio da una tricotomia ed una dicotomia:





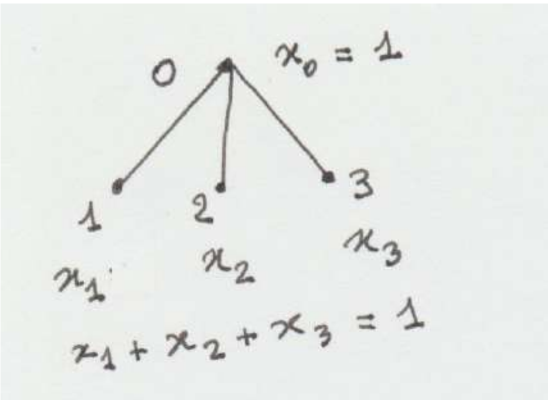
[26]

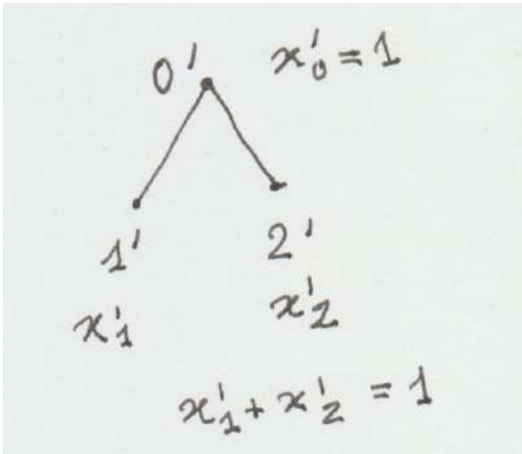
e di volerle tra loro "incollare" portando O' in 1:



[27]

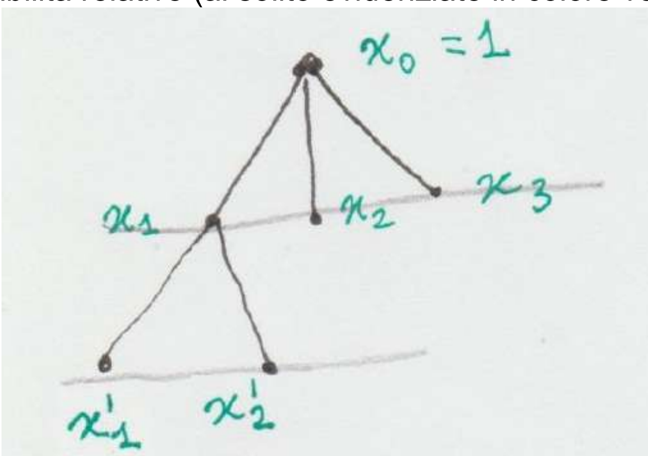
Supponiamo adesso che le k-tomie di partenza (k=3 o 2) diventino probabilistiche, ciascuna delle due con una propria distribuzione di probabilità (relative o totali non fa adesso nessuna differenza):





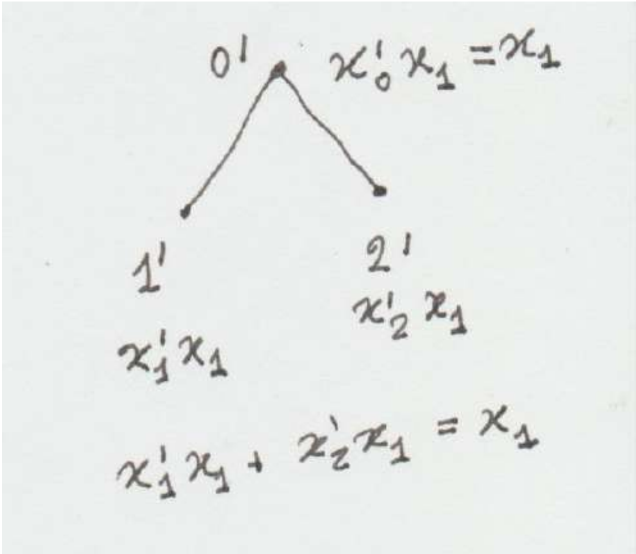
[28]

L'incollamento di cui alla figura [27] conduce al seguente grafo (non regolare) con le sue naturali probabilità relative (al solito evidenziate in colore verde):



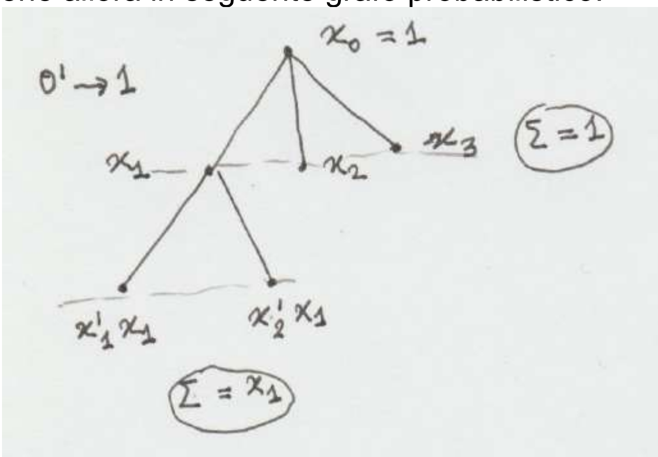
[29]

E' altrettanto naturale però voler incollare i due grafi in modo tale che O', la radice del grafo di destra che vogliamo incollare sul nodo 1 del grafo di sinistra, abbia la stessa probabilità di 1, cioè x_1, e non 1. Come si può fare? Basta modificare la distribuzione di probabilità sulla 2-tomia agendo in modo **proporzionale** su di essa, ossia moltiplicando tutte le probabilità dianzi assegnate per un fattore tale che il **nuovo** x'_0 sia uguale a x_1 e non a 1. In altre parole, basterà moltiplicare tutte le probabilità della dicotomia per x_1, ottenendo quindi:



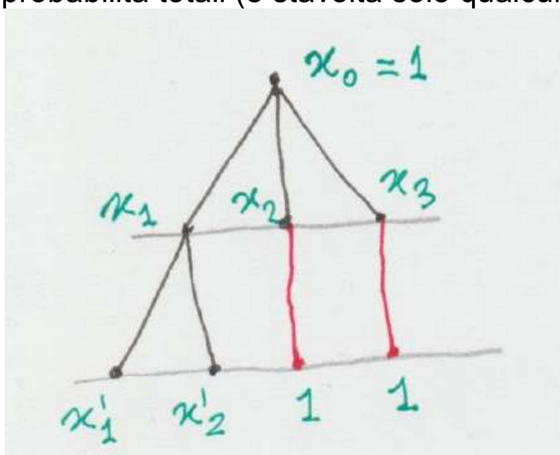
[30]

(si noti che tutte queste nuove probabilità sono ancora minori o uguali di 1)
e poi tranquillamente incollare O' in 1, nodi che hanno adesso la medesima probabilità. Si ottiene allora in seguente grafo probabilistico:

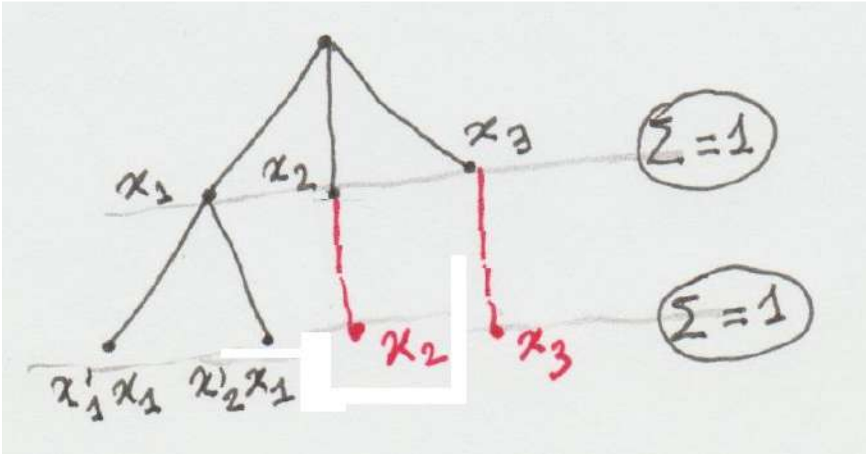


[31]

in cui evidentemente sono riportate delle probabilità totali (per le quali non abbiamo però stavolta utilizzato il colore rosso).
Concludiamo la lunga ma semplice Nota mostrando i grafi che si ottengono dai precedenti attraverso il procedimento di regolarizzazione, prima con le probabilità relative, e poi con le probabilità totali (e stavolta solo qualcuna di esse in rosso!):



[32]



[33]

Nota 8 (avanzata) - La probabilità come funzione in realtà di **due** variabili (nodi), una questione che rivedremo comparire sotto altra forma nel prossimo capitolo. Data infatti una distribuzione di probabilità relative su Γ, $P_\rho : \Gamma \to I_Q$, abbiamo già sostanzialmente proposto di utilizzare in luogo del simbolo $P_\rho(\mu)$, il seguente:
$P_\rho(\mu) = P(\mu,\mu\text{-}1)$, che mette in maggiore evidenza la dipendenza di $P_\rho(\mu)$ da μ-1 (oltre che ovviamente da μ). Possiamo però proseguire su questa strada, definendo per esempio P(μ,μ-2) come il prodotto P(μ,μ-1)P(μ-1,μ-2), ossia porre:
$P(\mu,\mu\text{-}2) = P_\rho(\mu)P_\rho(\mu\text{-}1)$, *etc.*, fino ad avere quindi una precisa definizione di una funzione P(μ,μ') definita per ogni coppia ordinata di nodi μ, μ' di Γ tali che μ' ≤ μ. Posto allora P(μ,μ') = 0 se μ' > μ, oppure se i due nodi sono tra loro **inconfrontabili** in Γ (ossia, se appartengono a cammini orientati diversi tra gli n = complessità dell'albero che vanno dalla radice di Γ ad uno degli n elementi massimali di Γ), otteniamo in definitiva una funzione $P : \Gamma x\Gamma \to I_Q$ (ΓxΓ = **quadrato cartesiano** di Γ, ovvero l'insieme delle coppie ordinate di elementi di Γ), funzione che soddisfa alle seguenti per noi interessanti uguaglianze:
P(μ,μ) = 1 per ∀ nodo μ di Γ, $P(\mu,\mu\text{-}1) = P_\rho(\mu)$, $P(\mu,R) = P_\tau(\mu)$ (R=radice di Γ).
P "contiene" cioè in sé sia P_ρ che P_τ, sicché, invece che descrivere prima queste due funzioni e poi P, si potrebbe seguire il cammino inverso, descrivere cioè prima P, e poi dedurre da essa sia P_ρ che P_τ, *sed de hoc satis*.

Cap. 3 - Un esempio universale, l'interpretazione insiemistica della probabilità (= statistica), probabilità condizionate e probabilità relative

Illustriamo adesso un esempio universale che spiega secondo noi tutto ciò in cui consiste il calcolo delle probabilità.

Prendiamo un insieme finito (non vuoto) A, il cui numero di elementi (che si dice **cardinalità** di A) indicheremo con il simbolo $|A|$, e consideriamo l'insieme S(A) dei suoi sottoinsiemi (un insieme anch'esso finito, di cardinalità $2^{|A|}$ (un insieme ricco di strutture algebriche sulle quali però in generale sorvoleremo). Introduciamo adesso la funzione

$P : S(A) \times S'(A) \to I_Q$, $P(X,Y) = |X \cap Y| / |Y|$

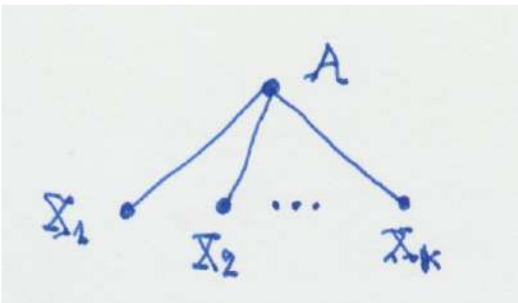
per ogni coppia ordinata di elementi di S(A) il cui secondo sia però diverso dall'insieme vuoto (per questo motivo abbiamo scritto S'(A) e non S(A), ma cerchiamo di non esagerare in pignoleria; ricordiamo che per l'insieme vuoto si usa il simbolo ∅, e che ∩ significa **intersezione**, corrispondente al simbolo logico ∧, che significa **congiunzione**, latino "et").

Interpretiamo P al seguente modo (che spiega anche la scelta della lettera P per designare tale funzione): P(X,Y) è la probabilità che un elemento x di X (si scrive x∈X, ∈ è il famoso **simbolo di Peano**) ha di essere anche un elemento di Y. P(X,Y) si scrive pure come P(X|Y), e si dice probabilità di X **condizionata** ad Y, o anche **relativa** ad Y (manifestamente: P(X|X) = 1). La probabilità condizionata P(X|A) = |X|/|A| di X relativa ad A si dice anche la probabilità **totale** di X, in simboli P(X) (ovviamente sottinteso, probabilità totale di **X in A**, ovvero di X in qualità di sottoinsieme di A). E' chiaro che la definizione è posta in modo tale che la cardinalità di un sottoinsieme X di A si ottenga da quella di A moltiplicandola per la probabilità di X in A:

|X| = P(X|A)*|A|

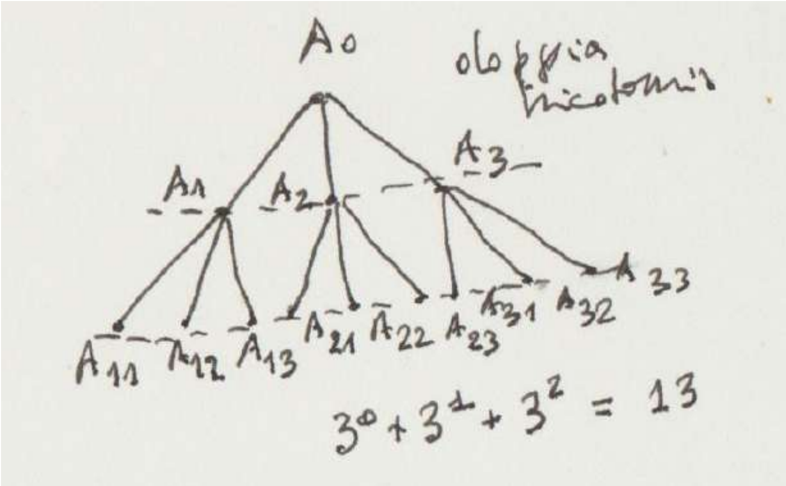
(per maggiore chiarezza, utilizzeremo qualche volta in questo capitolo il simbolo * come operatore di **moltiplicazione**).

E' chiaro che P soddisfa alla seguente condizione: se $X_1,...,X_k$ è una k-partizione di A, allora $P(X_1)+...+P(X_k) = 1$, ed è facile descrivere allora tale "situazione" mediante l'introduzione della seguente naturale k-tomia, e di tutta la nomenclatura dianzi stabilita:



[34]

Se ad essa aggiungiamo infatti l'interpretazione di uno dei "nostri" alberi, per esempio la seguente **doppia tricotomia**:



[35]

come di un dato insieme $A=A_0$ al livello 0, tripartito in A_1,A_2,A_3 al livello 1, e poi ciascuno degli A_i a sua volta tripartito al livello 2 (una sezione iniziale + tre sezioni successive), *etc.*, ecco che diventa più che manifesta l'analogia tra ciò che abbiamo illustrato in precedenza e questo particolare aspetto della teoria degli insiemi (partizioni, sottopartizioni, etc.), oltre che quella tra la funzione P dianzi introdotta e la funzione $P : \Gamma x \Gamma \rightarrow I_Q$ di cui alla Nota 8.

Nota 9 - Questa "vecchia" P, rispetto all'analoga "nuova" $P : S(A)xS'(A) \rightarrow I_Q$ che abbiamo poco fa introdotto, **sembra** avere più zeri, ma solo perché essa dà sempre risultato uguale a 0 a meno che non sia definita su una coppia di nodi che appartengono al medesimo "stelo" di Γ, ossia siano tra loro **confrontabili**, mentre P(X,Y) (con $X \in S(A)$, $Y \in S'(A)$) può per definizione non essere uguale a 0 anche se X, Y **non** sono tra loro confrontabili, ossia né $X \subseteq Y$ né $Y \subseteq X$ ($\subseteq$, o $\supseteq$, è il tradizionale simbolo di **inclusione insiemistica**). Insomma, i "nostri" grafi ad albero corrispondono idealmente unicamente a **catene** di sottoinsiemi di A, $A \supseteq X \supseteq Y$, *etc.*, procedendo come sempre dall'alto verso il basso, mentre ad ogni livello, in quella che potremo dire l'interpretazione insiemistica dell'albero, si ha a che fare solo con partizioni, ossia con sottoinsiemi che sono in particolare a due a due disgiunti.

Tutto ciò che precede ha la conseguenza che identità del tipo:

$P(\mu|\mu\text{-}1) = P_\rho(\mu) = P(\mu,\mu\text{-}1)$, $P(\mu|R) = P_\tau(\mu)$ (R = radice di Γ)

non appaiono più necessitare adesso di apposite spiegazioni, che in effetti non daremo anche se ci piace sottolineare che il passaggio dalle probabilità relative a quelle assolute (passaggio che ci interessava in maniera particolare) è adesso giustificato dalla seguente ovvia identità, per ogni catena $Z \subseteq Y \subseteq X$ di sottoinsiemi non vuoti di A:

$P(Z|X) = |Z|/|X| = (|Z|/|Y|)*(|Y|/|X|) = P(Z|Y)*P(Y|X)$

Non procediamo oltre su questa strada, ma affrontiamo prima di concludere il presente capitolo "insiemistico" (che si spera però non "insiemistificatore", come amava dire il famoso Prof. Bruno De Finetti, ordinario di Calcolo delle Probabilità a Roma al tempo degli studi dell'autore) una questione che diventerà poi la chiave per risolvere un punto essenziale nella discussione che ci aspetta nel prossimo capitolo. Una questione che è al contempo utile ed interessante, e per fortuna abbastanza semplice, ma forse non semplicissima per chi non sia abituato a certo linguaggio matematico, ancorché alquanto "elementare".

Per due sottoinsiemi non vuoti X, Y di A, calcoliamo la probabilità totale di $X \cap Y$ in A. Sarà evidentemente, come già sappiamo:

$P(X \cap Y) = P(X \cap Y|A) = |X \cap Y|/|A| = |X \cap Y|/|Y|*|Y|/|A| = P(X|Y)P(Y|A) = P(X|Y)P(Y)$

ed analogamente:

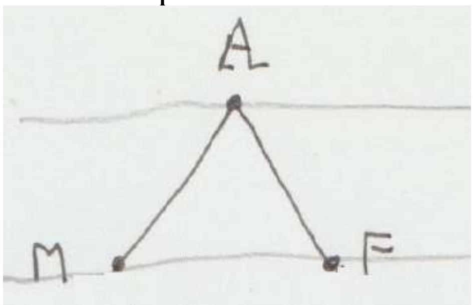
$P(Y \cap X) = P(Y \cap X|A) = |Y \cap X|/|A| = |Y \cap X|/|X|*|X|/|A| = P(Y|X)P(X|A) = P(Y|X)P(X)$

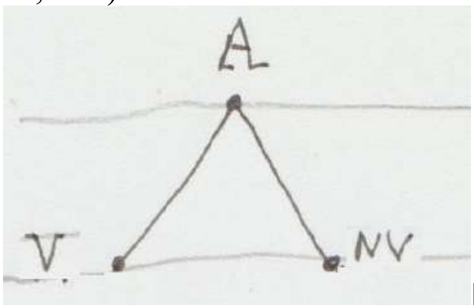
L'identità che abbiamo così ottenuto, e che evidenziamo con il grassetto:

P(X|Y)P(Y) = P(Y|X)P(X)

costituisce la famosa (ma in fondo elementare) **formula di Bayes**,

Mostriamo adesso con un istruttivo facile esempio come si possa fare ricorso alla formula in parola. Supponiamo che l'insieme A sia costituito da una popolazione nella quale siano presenti due caratteri indipendenti, diciamoli M e V (introducendo un esempio di triste attualità, M = "maschi", V = "vaccinati almeno una volta contro il Covid-19"), che inducono due distinte bipartizioni in A: $A = M \cup NM$, $A = V \cup NV$, che potremo descrivere graficamente così (pure un opportuno disegnino "geometrico" potrebbe aiutare a comprendere meglio la situazione, bipartendo per esempio un quadrato in due parti dapprima tramite una linea verticale, d'onde punti che si trovano alla sinistra della linea e punti che si trovano alla destra, e poi con una linea orizzontale, d'onde punti che si trovano al di sopra o al di sotto della linea, *etc.*):

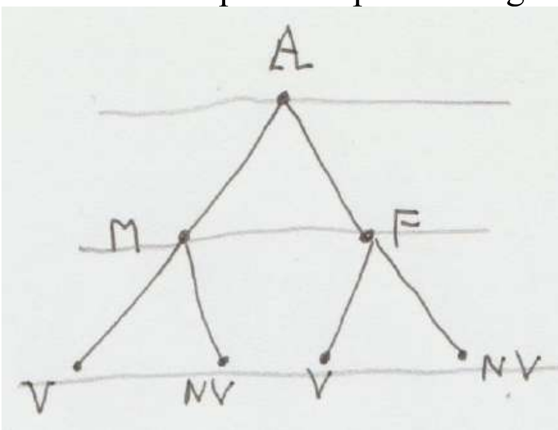




[36]

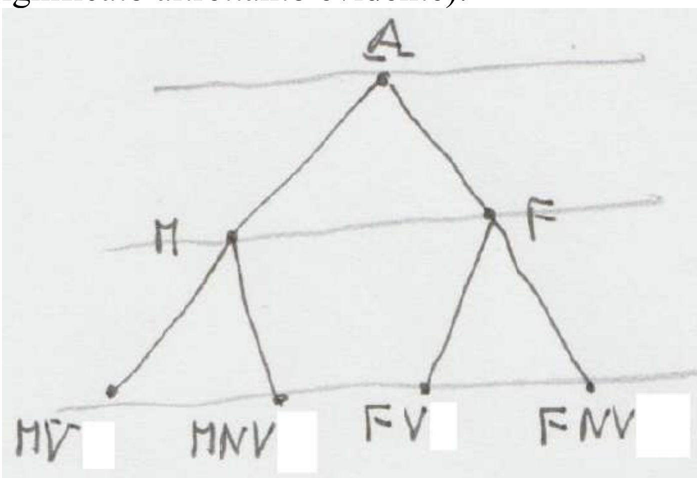
Leggiamo NM come "non-M", ossia "femmine", il che spiega l'introduzione della lettera F in luogo di NM (ignoriamo così volutamente certe recenti abominevoli tendenze ... ideologiche provenienti da oltreoceano, come del resto ogni altro simile pattume), e volendo NV come *no-vax*. Diciamo ancora che $\cup$ significa **unione**, corrispondente al simbolo logico $\vee$, che significa **disgiunzione inclusiva**, latino "vel". Inoltre, che NM si può stenograficamente indicare con $\neg$M, introducendo il simbolo logico di **negazione** $\neg$ (simboli peraltro già utilizzati e spiegati nella figura [3]).

Possiamo adesso ... complicare la situazione, volendo comparare tra loro le dette due bipartizioni, pervenendo così per esempio alla seguente **doppia dicotomia**:



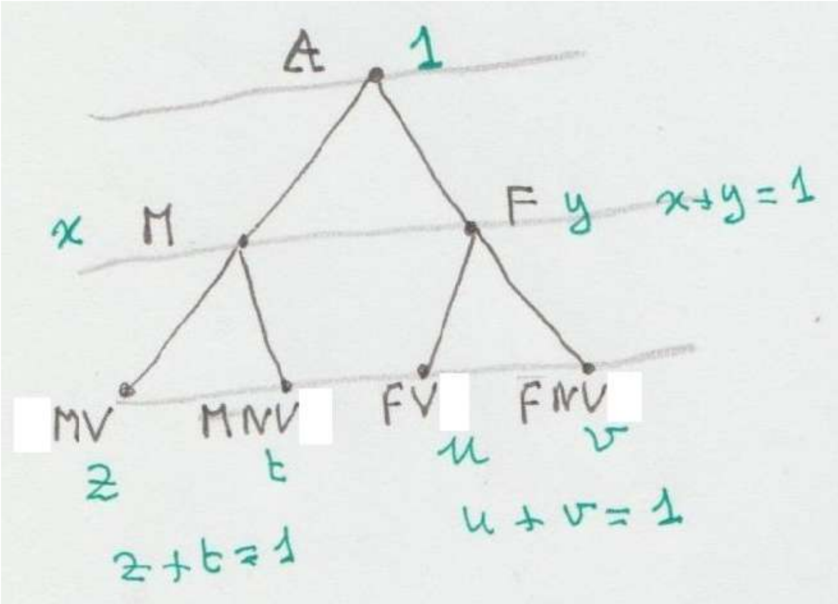
[37]

che ha un significato abbastanza chiaro in sé, sebbene si trovi lo stesso simbolo a designare due nodi **diversi**, sicché la modifichiamo subito nella maggiormente precisa (e dal significato altrettanto evidente):

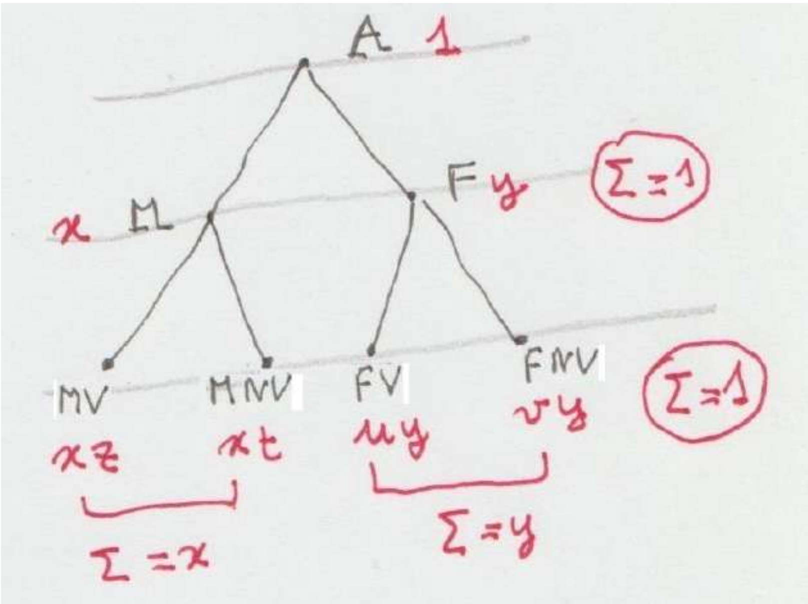


[38]

Riportiamo adesso qui di seguito la dicotomia con la sue probabilità, rispettivamente relative e totali, probabilità a noi sconosciute, donde l'uso di lettere in luogo di numeri (e saremmo in verità curiosi di conoscerli questi numeri, convinti che potrebbero forse insegnare qualcosa):

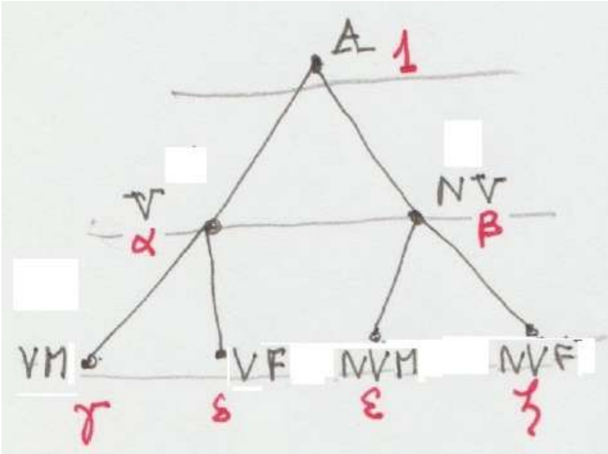


[39]



[40]

Ci poniamo adesso il seguente **problema**: come possiamo passare dal grafo probabilistico di cui in figura [40], all'analogo grafo dove però i due livelli M-F e V-NV sono tra loro **scambiati**? In altre parole, si possono (speriamo facilmente) determinare, ovviamente in funzione di x,y,z,..., le probabilità totali che competono al seguente grafo, un grafo che potremo suggestivamente designare come il grafo **inverso** del precedente? (probabilità sconosciute al pari delle x,y,z,..., che sono indicate adesso con le lettere greche α,β,γ,...).



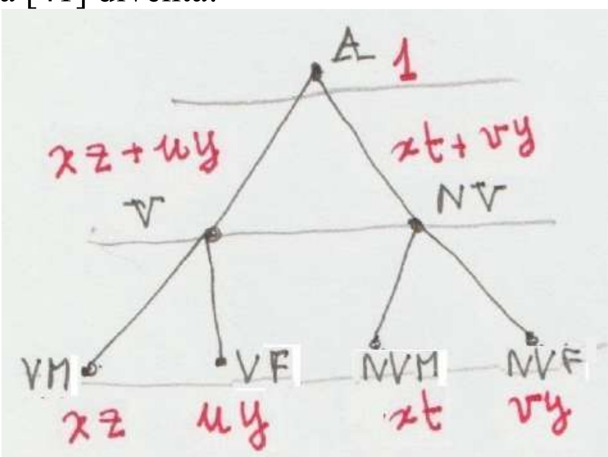
[41]

Per fortuna la risposta è abbastanza facile, poiché α vale evidentemente xz+uy, ossia la **somma** delle due occorrenze di V nel secondo livello del grafo [40], ed analogamente β = xt+vy (questo si chiama il **teorema della probabilità totali**, che è del tutto evidente in un contesto insiemistico quale quello qui in considerazione; torniamo a dire che sarebbe forse proficuo per fini didattici introdurre un bel disegnino geometrico come poco fa suggerito, in cui andrebbero messe in evidenza, per esempio ricorrendo a colorazioni diverse, le sovrapposizioni tra i sottoinsiemi che costituiscono le due bipartizioni, cioè le loro intersezioni). Cosa si può dire invece di γ, di δ, *etc.*? E' chiaro che, in conformità con l'analisi e i simboli precedenti, dalla formula di Bayes si possono dedurre le seguenti identità:

$\gamma = P_\tau(MV) = P_\rho(MV)P_\tau(V) = P(M|V)P_\tau(V) = P(V|M)P_\tau(M) = P_\tau(VM) = xz$

$\delta = P_\tau(FV) = P_\rho(FV)P_\tau(V) = P(F|V)P_\tau(V) = P(V|F)P_\tau(F) = P_\tau(VF) = uy$, *etc.*.

Insomma, la figura [41] diventa:



[42]

ed è "divertente" notare che davvero, come deve essere:

$\alpha + \beta = (xz+uy) + (xt+vy) = (xz+xt) + (uy+vy) = x + y = 1$
$\gamma + \delta = xz + uy = \alpha$
$\varepsilon + \zeta = xt + vy = \beta$

e ancora (infine):

$(\gamma+\delta) + (\varepsilon+\zeta) = (xz+uy) + (xt+vy) = 1.$

Vale a dire, tutto torna, e ciò ... ci riempie di soddisfazione.

Chiudiamo il capitolo proponendo al lettore che voglia ulteriormente "divertirsi" di verificare che davvero, "invertendo" adesso tra loro il primo e il secondo livello presenti nel grafo [42], si ritrova il grafo di partenza [40], sebbene sembri apparentemente che l'inversione implichi un certo aumento della complessità formale del grafo probabilistico d'arrivo rispetto a quello di partenza.

Cap. 4 - Il paradosso di Monty Hall

Dopo la precedente lunga ma abbastanza facile "premessa", veniamo alla sua annunciata applicazione all'argomento che costituiva il nostro principale interesse quando abbiamo iniziato ad occuparci di didattica del Calcolo delle Probabilità. Interesse in generale nei confronti dei cosiddetti "paradossi", una discussione di alcuni dei quali maggiormente noti in ambito matematico si trova nella Parte III del lavoro citato solo attraverso il relativo URL nel Cap. 0, "Breve presentazione (critica) del teorema di incompletezza di Goedel", Detta parte si apre con la seguente epigrafe: «Sciocchezze quali la maggior parte dei paradossi non saranno mai minimamente prese in considerazione dalla vera logica» (Carl Prantl, *Geschichte der Logik im Abendlande*, 1855, vol. I, p. 95), che la dice lunga sulla nostra opinione al riguardo.

Poiché si tratta di una questione piuttosto nota che, ripetiamo, vogliamo qui affrontare come semplice "esercizio" nell'ambito dell'approccio teorico dianzi descritto, ci limiteremo ad introdurla attraverso una semplice citazione, senza indulgere in riferimenti di carattere storico, o "psicologico", che sarebbero comunque interessanti, ma che lasciamo ad un esteso lavoro in preparazione da parte del caro amico filosofo RVM, filosofo nel senso più ampio del termine:

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, ***who knows what's behind the doors****, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?*
(https://en.wikipedia.org/wiki/Monty_Hall_problem)

Abbiamo evidenziato con il grassetto un punto importante del problema, un punto che in analisi frettolose e superficiali della questione non viene ben evidenziato, e non lo è nemmeno adesso nella riportata citazione. Non basta infatti introdurre la dicotomia logica "il conduttore **SA** oppure **NON SA**", bisogna procedere con uno di quegli alberi che abbiamo definito a cascata di dicotomie, almeno nella sua parte iniziale. Se il conduttore NON SA la cosa finisce lì, se invece SA, allora bisognerà chiedersi se utilizzerà o non utilizzerà questa sua conoscenza, lasciando per esempio la decisione di aprire una porta in luogo di un'altra al caso (utilizzando una "ruota della fortuna", lanciando un dado, facendo scegliere al concorrente stesso, il quale ovviamente **non sa** cosa ci sia dietro le porte, *etc.*). Se non la utilizzerà la cosa finisce lì, se invece la utilizzerà bisognerà chiedersi con **quale finalità** la utilizzerà. Per aiutare il concorrente nella sua speranza di vittoria qualora abbia sbagliato la scelta iniziale? Per indurlo viceversa in errore qualora al contrario l'abbia azzeccata, e risparmiare quindi sul costo della trasmissione - mancata concessione del premio? Soltanto per allungare il brodo dello spettacolo aumentando la *suspense* negli spettatori? Un po' di qualcuno di codesti motivi, un po' dell'altro, a seconda del momento e delle circostanze?

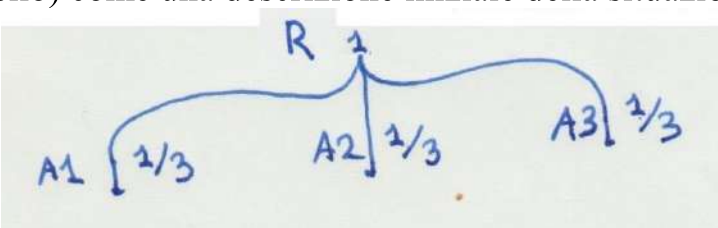
Bene, diamo connotati più precisi al problema che intendiamo risolvere esplicitando l'ipotesi usuale che il conduttore sappia, e che utilizzi la sua conoscenza esclusivamente a fini di spettacolo. Vale a dire, che sia dagli organizzatori sin dall'inizio previsto che, dopo la scelta iniziale effettuata dal concorrente, venga aperta una delle porte dietro le quali c'è una capra. In altre più precise parole, che il conduttore agirà, sia **costretto** ad agire, rispettando **entrambe** queste regole:

1 - non farà mai aprire la porta scelta dal concorrente;
2 - non farà mai aprire la porta con il premio (la quale potrebbe coincidere peraltro con quella scelta dal concorrente, nel qual caso le due regole si ridurrebbero ad una sola).

Sufficiente? Non proprio, poiché le domande di Wikipedia sono rivolte al concorrente, il quale deve essere lui a calcolare le probabilità di vittoria in un caso (non modificando la propria scelta iniziale) o nell'altro (invece modificandola), e quindi deve essere pure lui messo al corrente - almeno in parte - delle regole del gioco, altrimenti potrebbe per esempio legittimamente ritenere che la proposta di cambiare scelta gli venga rivolta soltanto quando avrebbe vinto con quella iniziale, *etc. etc.*. Insomma, la dicotomia **SA** oppure **NON SA** va estesa anche all'altro partecipante allo spettacolo, non solo all'*host*, ovviamente una conoscenza con modalità diverse a seconda dei ruoli (conduttore ... onnisciente, concorrente che non saprà fino alla fine dietro quale porta sia il premio, ma che sa sin dall'inizio come si svolgerà il gioco, altrimenti non potrà in nessun modo prendere quella decisione univoca "razionale" che si pretende da lui).

Bene, ciò premesso affrontiamo dunque il problema come un'applicazione diretta di quanto illustrato nei capitoli precedenti, nella convinzione che, se la logica dell'intelletto o l'intuizione possono a volte ingannarsi, il risultato di un calcolo ben condotto (algoritmo) non può mai indurre in errore. Cercheremo pertanto di ridurre lo spazio della pura argomentazione verbale, nonostante siamo altrettanto persuasi che, se ben condotti, siffatti ragionamenti possono condurre a spiegazioni altrettanto soddisfacenti di quella matematica che qui conseguiremo.

Tanto per cominciare, si potrebbe semplicemente introdurre la seguente tricotomia probabilistica (con R indicheremo d'ora in avanti la radice dell'albero, o ... dell'alberello) come una descrizione iniziale della situazione in esame:



[43]

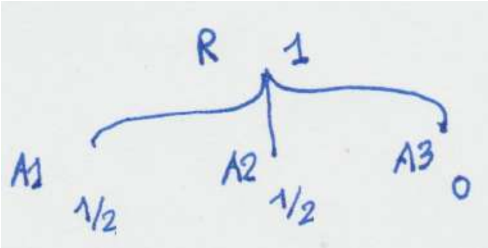
A1, A2, A3 significano ovviamente l'auto si trova dietro la porta 1, *etc.*, le tre modalità sono ovviamente mutuamente escludentisi ed assolutamente equiprobabili. Indicheremo analogamente con S1, S2, S3 le tre possibili scelte del concorrente, e con P1, P2, P3 le tre possibili aperture delle porte (indicheremo con lo stesso simbolo anche le porte) da parte del conduttore. Notiamo che la scelta di questa lettera iniziale maiuscola P per "porta" ci porterà a scrivere in questo capitolo le probabilità con il simbolo Pr. Ciò fatto, porremo la domanda fondamentale (DF):

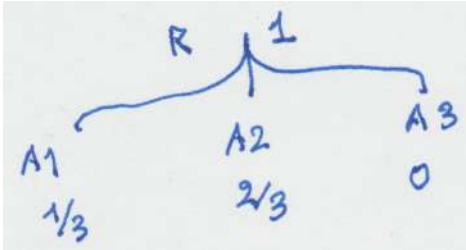
DF: Come deve essere da parte del concorrente razionalmente modificata la precedente distribuzione di probabilità dopo che, avendo egli scelto per esempio S1, il conduttore abbia aperto la porta P3 mostrando la capra che ci sta dietro?

(esageriamo ricordando ancora una volta che abbiamo convenuto che il conduttore non può aprire P1 dopo che è stata effettuata la scelta S1, e che in ogni caso non può aprire la porta con il premio?).

In altre parole, come il concorrente deve modificare la distribuzione iniziale di equiprobabilità allorché il suo stato dell'informazione è mutato rispetto a quello che aveva inizialmente, dal momento che viene a sapere, **dopo** aver effettuato la scelta S1, che Pr(A3) è uguale a 0, e non ad 1/3? Sottolineiamo che avremo sempre a che fare qui con quelle che abbiamo definito probabilità a priori effettive, e quindi **oggettive**. Queste potrebbero diventare soggettive solo in conseguenza di qualche "soffiata", o di qualche accordo con i produttori del programma, nel qual caso cambierebbe ovviamente lo stato dell'informazione almeno di qualche "osservatore" che venga messo al corrente di tali conoscenze ... suppletive!

L'alternativa cade (naturalmente) soltanto su una delle due seguenti tricotomie probabilistiche:





[44]

delle quali è quella di destra ad essere razionale (**con le regole 1 e 2 che abbiamo dianzi enunciato**), e non quella di sinistra, anche se appunto molti cadono nell'errore di considerarla tale (d'onde il termine introdotto di "paradosso").

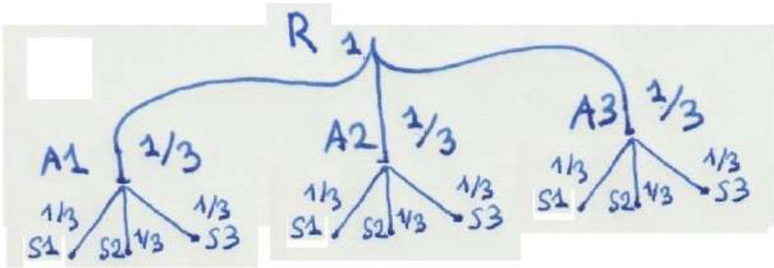
Nota 11 - In relazione a quanto esposto nella Nota 1, la QI del grafo di sinistra è uguale a 1/4, mentre quella del grafo di destra è uguale a 1/3, quindi **maggiore** di 1/4 (all'inizio del gioco la QI del concorrente era ovviamente 0, quella del conduttore è per le attuali ipotesi sempre uguale ad 1). Il "rischio" per il concorrente è ritenere di essere in possesso di una QI **inferiore** a quella che egli invece possiede effettivamente dopo l'apertura della porta 3 (apertura che, ricordiamo, è stata condizionata dalla sua scelta S1, fino al caso che il conduttore non abbia avuto altra possibilità che P3, ossia sia stato **costretto** ad aprire la porta 3), Ovvero, di **non** saper sfruttare appieno la sua QI, e per conseguenza di ritenere erroneamente "giusto" il grafo di sinistra!

Per convincere coloro che esprimono un giudizio errato (o forse meglio, imperfetto, qualora non abbiano ben chiare le regole del gioco e la loro rilevanza), si cerca di argomentare che, date le condizioni con le quali si è sviluppata la situazione, la Pr(A1) non può essere cambiata dall'1/3 iniziale, e che quindi, a parte Pr(A3) che è diventata addirittura 0, quella che è cambiata (è stata ... costretta a cambiare) è unicamente Pr(A2), che dovrà quindi essere posta uguale a 2/3 (ossia, la probabilità iniziale che il premio fosse dietro ad una qualsiasi delle due porte P2 e P3), poiché la somma delle tre probabilità deve comunque essere sempre uguale ad 1.

Raccogliamo la conclusione in un (piccolo):

Teorema - Ferme restando le regole 1 e 2, effettuata la scelta S1, aperta la porta P3, se si cambia la scelta iniziale (quindi da S1 a S2) la probabilità di vittoria raddoppia, da 1/3 a 2/3.

Con codesto "puro ragionamento" si potrebbe chiudere la discussione, ma noi vogliamo andare avanti utilizzando la nostra prediletta matematica. Cominciamo così con l'introdurre, in luogo della tricotomia probabilistica di cui alla figura [43], una 9-tomia probabilistica la quale, **diluita** come abbiamo nei capitoli precedenti spiegato, ammette la seguente forma:

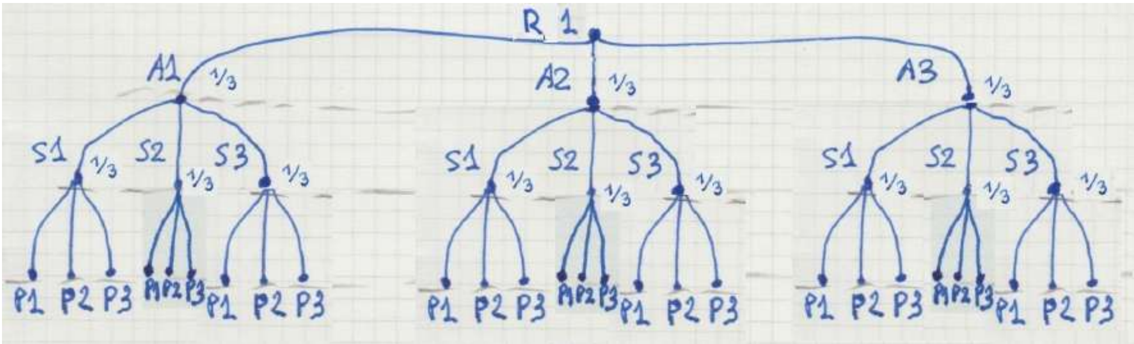


[45]

Siamo davanti a un grafo uniforme di caratteristica 3 ed altezza 2, in cui sono state riportate (sebbene non facendo ricorso stavolta ai circoletti e al colore verde) le **probabilità relative** inerenti alla situazione in oggetto (ai sensi del Cap. 2), ed in cui ci siamo permessi un abuso di notazione: ossia abbiamo indicato sempre con lo stesso simbolo nodi **diversi**. Un abuso che evita però appesantimenti eccessivi della trattazione, e che non dovrebbe comunque indurre in ambiguità, dal momento che, se si scrive per esempio Pr(S3|A2), si capisce bene che si sta parlando dell'S3 che appare a destra nella terza sezione del grafo (la seconda delle 3 presenti al II livello), quello che sta "sotto" ad A2, e non tanto per dire quello che appare a destra nella quarta sezione del grafo (la terza del II livello), ossia il nodo che sta "sotto" ad A3.

Ci si potrebbe chiedere: dove sta la 9-tomia? Sta nei 9 cammini orientati che da R conducono ai 9 elementi massimali dell'albero, cammini che potremo indicare con i simboli: R-A1-S1, R-A1-S2, R-A1-S3, *etc.*, sempre leggendo dall'alto verso il basso e da sinistra verso destra. Individuato l'algoritmo, è facile elencarli tutti in maniera completa, ovvero non è possibile saltarne qualcuno, e si trova che ciascuno dei 9 ha ovviamente una probabilità (che è adesso **totale**) uguale a 1/9=(1/3)*(1/3). Bene, è chiaro che si vince con R-A1-S1, R-A2-S2, R-A3-S3, cioè 1/9+1/9+1/9=1/3, mentre si perde con tutte le altre "combinazioni", ossia 6/9=2/3. Vale a dire, si ritrova banalmente quanto già si sapeva, cioè il valore delle probabilità assolute concernenti al gioco **prima** però dell'apertura della porta P3, e della proposta di cambiare la scelta inizialmente effettuata.

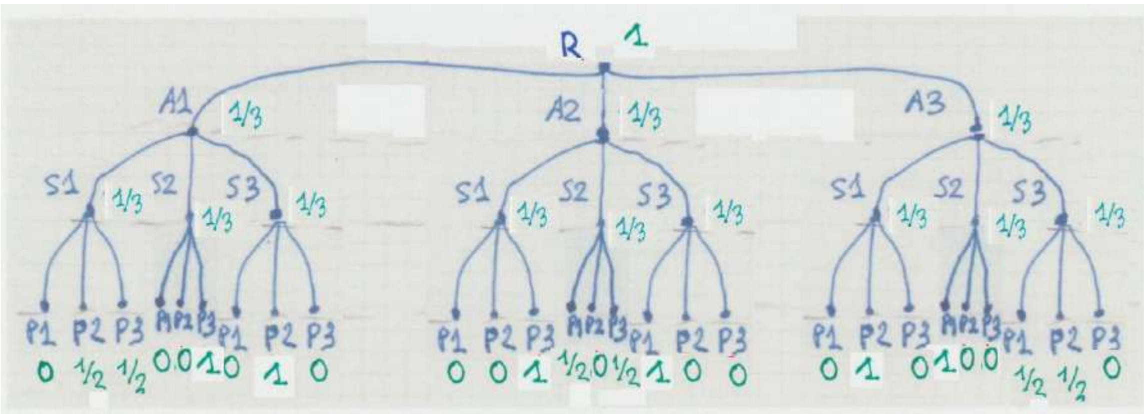
Come possiamo prendere in considerazione codesto ulteriore elemento (che in effetti **cambia** la situazione iniziale e la fa diventare un'altra)? Ma semplicemente introducendo il seguente albero uniforme di caratteristica 3 e di altezza però uguale a 3 in luogo di 2, quindi una 27-tomia diluita (i 3 livelli di diluizione vengono presentati in un ordine che potremmo dire **crono-logico**, ossia prima si sistema il premio dietro qualche porta, poi il concorrente effettua la sua scelta, infine il conduttore decide quale porta aprire):



[46]

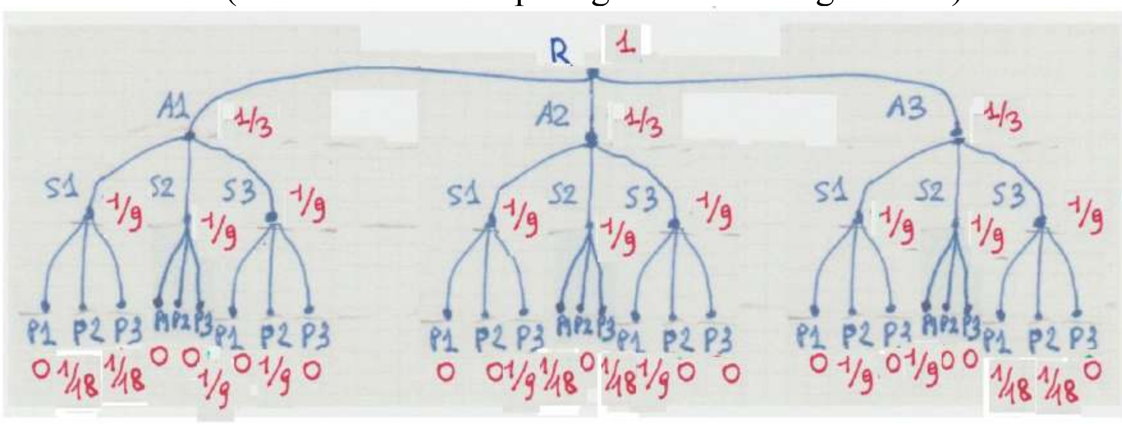
La 27-tomia è naturalmente rappresentata dai 27 cammini orientati che si deducono dal grafo precedente, ossia R-A1-S1-P1, R-A1-S1-P2, *etc.*, anche qui tutti elencabili **facilmente** in un certo ordine e in maniera completa, senza rischiare omissioni. Un grafo di complessità (3^0=1,3^1=3,3^2=9,3^3=27), quindi con $3^0+3^1+3^2+3^3=(3^4-1)/(3-1)=40$ vertici e 39 lati (la prima uguaglianza si deduce dalla seguente "bella" e semplice identità polinomiale, immediatamente verificabile moltiplicando il denominatore che sta a sinistra per il polinomio che sta a destra: $(x^{n+1}-1)/(x-1) = 1+x+x^2+...+x^n$).

Il grafo precedente non è un grafo probabilistico, poiché manca ogni indicazione concernente una distribuzione di probabilità Pr, ma è semplice introdurla (e, ripetiamo, probabilità a priori effettiva) tenendo conto di tutte le **regole** dianzi formulate:



[47]

probabilità relative (utilizzando di nuovo per esse il colore verde)
(la somma delle Pr per ogni sezione è uguale a 1)



[48]

probabilità totali (utilizzando di nuovo per esse il colore rosso)
(la somma delle Pr per ogni livello è uguale a 1, e per ogni sezione
di una data radice è uguale alla probabilità totale di detta radice)

Nota 12 - Se il conduttore utilizza altre regole per l'apertura delle porte, per esempio ne apre una a caso scelta tra tutte e 3, oppure ne apre una a caso scelta tra le 2 diverse da quella scelta del concorrente, cambiano ovviamente le probabilità relative e totali riportate al III livello dei grafi precedenti, ma ovviamente anche in queste varianti sarà facile calcolare ogni tipo di probabilità (a priori effettiva, relativa o totale) che si possa desiderare.

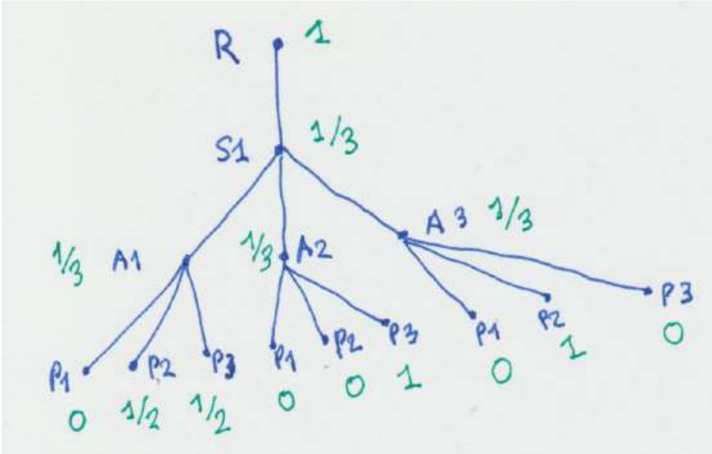
Bene, dopo aver insistito sulla circostanza che il grafo completo precedente è facile da descrivere (automatismo), e che in maniera altrettanto automatica lo abbiamo dotato di Pr relative (traducendo le "regole" del gioco), passando poi ancora facilmente, ossia automaticamente, da queste alle corrispondenti Pr totali, possiamo chiederci: siamo soddisfatti di aver ... complicato la situazione fino a tal punto?

Purtroppo no, se vogliamo per esempio dare precisa risposta alla domanda fondamentale DF. Abbiamo infatti la sensazione di essere sulla buona strada, ma il grafo di cui sopra fornisce le Pr di cammini del tipo R-A-S-P, e non di cammini quali R-S-P-A, cioè di quelli che invece servirebbe analizzare al fine di determinare la corretta risposta alla DF. Insomma, siamo chiamati a dover **invertire** l'ordine dal quale avevamo preso le mosse, e ciò realizzeremo passando:

1 - prima da R-A-S-P a R-S-A-P, facile, perché si tratta di livelli del tutto **indipendenti** tra loro (in termini più precisi, i due grafi probabilistici di cui sopra rimangono evidentemente "invariati" se si scambiano le A con le S);

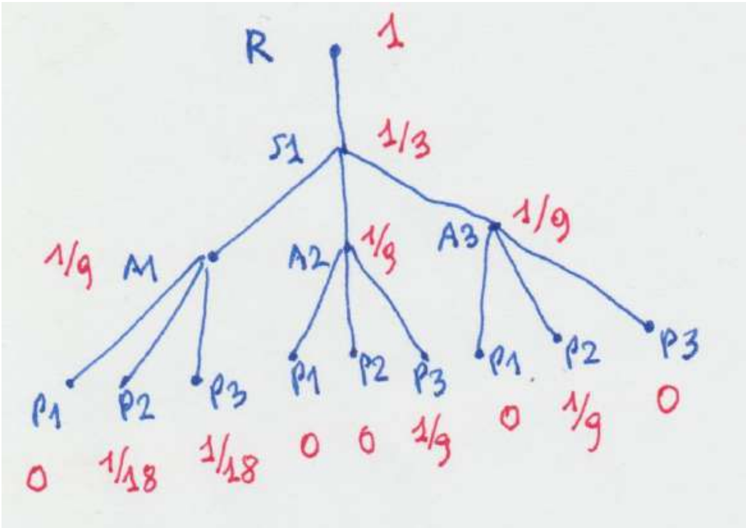
2 - in seguito da R-S-A-P a R-S-P-A, passo per il quale il lettore che è arrivato pazientemente a leggerci fin qui avrà già compreso che faremo ricorso alla formula di Bayes, cosa non facilissima ma comunque un automatismo di quelli che ci piacciono. A questo punto osserviamo però esplicitamente che mentre appare lecito cambiare l'ordine tra A e P, non sarebbe invece lecito proporre uno scambio dei due livelli S e P, per motivi crono-logici che sono anche **causali**, al riguardo si veda pure la successiva Nota 13.

Per quanto riguarda il passo 1, senza stare a riscrivere completamente gli alberi di nostro interesse con le A al posto delle S e viceversa, e tenuto conto che, al fine di rispondere alla DF è sufficiente prendere in considerazione i cammini che cominciano con R-S1, ci rendiamo subito conto che basteranno i seguenti disegni, che rappresentano ovviamente solo **un terzo** dei grafi probabilistici completi ai quali appartengono:



[49]

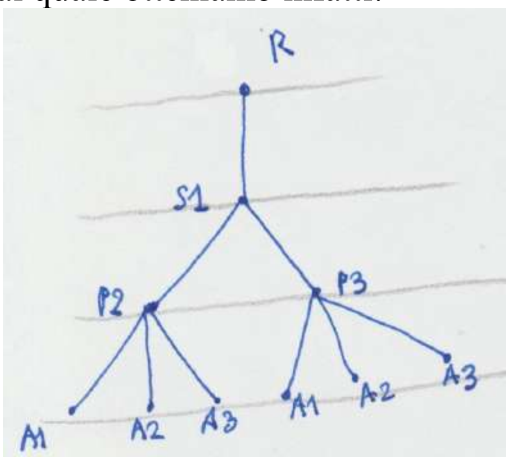
probabilità relative



[50]

probabilità totali

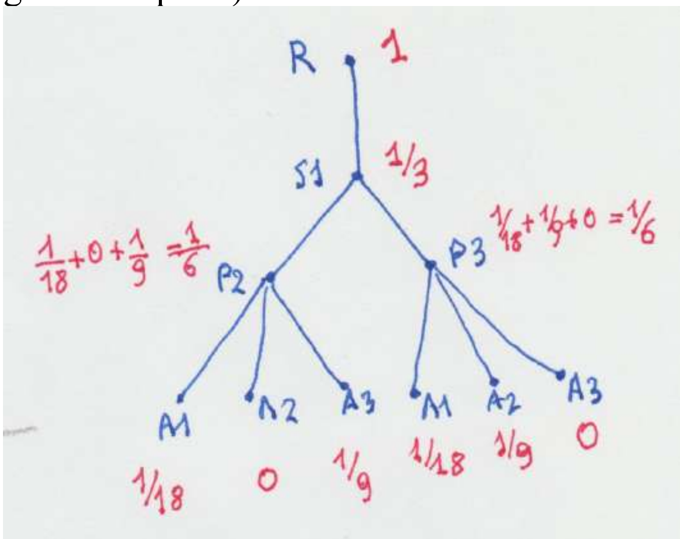
Quello che dobbiamo invertire (per quanto concerne il II e III livello) è quest'ultimo grafo, dal quale otteniamo infatti:



[51]

Un grafo che abbiamo ulteriormente semplificato eliminando del tutto il riferimento alla porta P1, dal momento che se è stata effettuata la scelta S1, il conduttore rispetterà la regola di non aprire la porta scelta dal concorrente.

Mancano nel precedente grafo [51] le probabilità totali che ci interessano, ed eccole qui di seguito riportate (ricordiamo che si tratta solo di **un terzo** del corrispondente grafo completo):



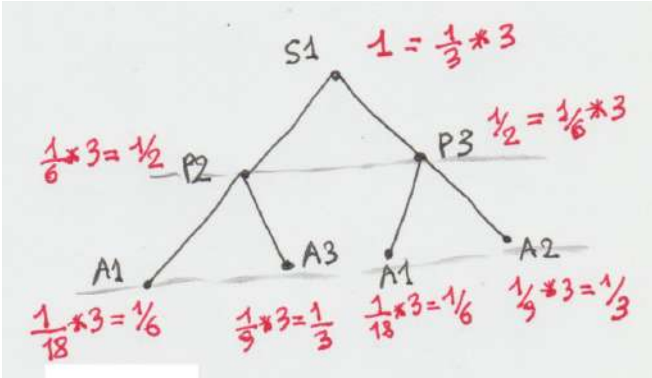
[52]

Esse sono state immediatamente desunte da un'applicazione della formula di Bayes e dal teorema delle probabilità totali, come abbiamo spiegato nella parte finale del Cap. 3.

Nota 13 (lunga, ma importante) - La probabilità totale dell'A3 che sta "sotto" a P2 nel grafo [52] è uguale alla probabilità totale del P2 che sta "sotto" ad A3 nel grafo che si vuole "invertire", ossia [50], *etc.*. Bisogna in verità riconoscere che, in un ciclo di lezioni sui fondamenti del Calcolo delle Probabilità rivolto a dei giovani che non siano tutti aspiranti specialisti, sarebbe doveroso fornire almeno qualche indicazione sul come e perché la formula di Bayes, certamente valida in un ben preciso ambito insiemistico, si possa applicare anche a diversi contesti, quali quelli che possono essere descritti da un grafo probabilistico qualunque. E ciò tenuto viepiù conto del fatto che non sempre in concreto i livelli dei nostri grafi si possono tra loro scambiare, per motivi sia logici che crono-logici: di sicuro non si potrà farlo quando l'aspetto cronologico implichi nessi **causali** tra le varie modalità e sotto-modalità in esame, un'indebita inversione potrebbe condurre a quelli che ci piace definire come **"paradossi alla Terminator**" (viaggi nel tempo e sciocchezze simili, che non ci interessano nemmeno come lettori di fantascienza - che è cosa ben diversa dal ... fantaridicolo) Tanto per dire, come abbiamo già detto, nel caso che è precipuo oggetto del nostro attuale interesse, non avrebbe senso voler scambiare il livello delle S con quello delle P, dal momento che la scelta di quale porta aprire **dipende** causalmente dalla scelta preliminare effettuata dal concorrente. Ciò premesso, accenniamo soltanto alla circostanza che di un grafo probabilistico quale quello riportato nella figura [40] si può sempre facilmente costruire un "modello" insiemistico (a prescindere dal fatto che quel grafo in particolare era già esso stesso un modello insiemistico), semplicemente introducendo prima un comune multiplo D di tutti i denominatori presenti in qualcuna delle rappresentazioni delle frazioni x, y, xz,..., e poi un insieme di cardinalità D, sul quale verrà definita una bipartizione in due sottoinsiemi di cardinalità Dx, Dy, indi bipartendo adesso il primo di questi due sottoinsiemi in due sottoinsiemi di cardinalità Dxz, Dxt, *etc.*. Insomma, una situazione

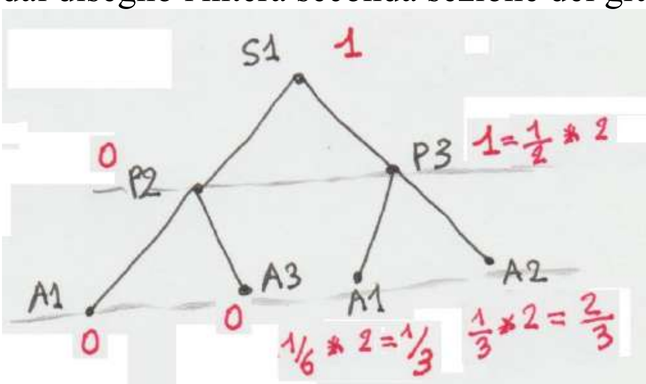
si potrà sempre descrivere in maniera equivalente con un'interpretazione insiemistica, e sempre in ultima sostanza, nell'individuazione e nella gestione di una distribuzione di probabilità, si tratta di effettuare dei semplici calcoli aritmetici. Quelli che abbiamo cercato fin qui di illustrare sono soltanto dei metodi, degli ausili per l'intelletto, escogitati al duplice scopo di rendere più agevoli tali calcoli, e di minimizzare la possibilità di errore (non diciamo probabilità, se vogliamo evitare di correre il rischio di precipitare in ... un giro vizioso, senza essere mai più capaci di tirarcene fuori!). Tanto per dire ancora, una sorta di "modello canonico" per il nostro grafo [46] si può evidentemente costruire come segue. Si comincia con l'introdurre l'insieme, chiamiamolo L (dall'iniziale di "lettere"), costituito da tutti i simboli usati, L = {A1,A2,A3,S1,..,P1,...}. L è un insieme di 9 elementi che ammette una naturale 3-partizione **uniforme** (ossia del tipo 3,3,3): L = A$\cup$S$\cup$P, dove abbiamo posto naturalmente A = {A1,A2,A3} *etc.*. Indi si considerano prima L^3, l'insieme delle terne ordinate di elementi di L, che sono in numero di 9^3 = 729, e poi all'interno di L^3 il sottoinsieme costituito da tutte quelle terne che sono prive di ripetizioni, un insieme di cardinalità uguale a 9*8*7 = 504. Procedendo oltre, si va a considerare all'interno di questo sottoinsieme il sotto-sottoinsieme costituito da quelle terne che hanno un solo elemento in A, uno solo in S, e uno solo in P, un insieme che ha cardinalità 9*6*3 = 162. Otteniamo così l'insieme, diciamolo C, corrispondente a **tutti** i possibili "cammini orientati" oggetto della nostra attenzione, del tipo cioè per esempio A-P-S, *etc.*, il quale a sua volta ammette una 3-partizione uniforme (del tipo 54,54,54): L = CA$\cup$CS$\cup$CP, dove CA significherà palesemente l'insieme di tutti quei cammini che cominciano con un elemento di A, *etc.*. In CA (allo stesso modo che in CS e in CP) si considera poi la bipartizione uniforme (del tipo 27,27) CA = CASP$\cup$CAPS, ed ecco che abbiamo per esempio trovato un insieme, ossia CASP, che corrisponde esattamente a tutti i cammini orientati del grafo [46] (e, giunti a questo punto, sottolineiamo che CASP ci interessa, allo stesso modo che CSAP e CSPA, mentre non ci interessano, per le dianzi esposte ragioni crono-logiche e causali, né CPAS o CPSA, oppure ancora CAPS). C'è davvero bisogno di dire adesso che il nodo A1 è "rappresentato" dai sottoinsieme di CASP costituito da tutti quei cammini che hanno come primo elemento A1 *etc.*,, e che siamo finalmente di fronte ad una 3-partizione uniforme di CASP (del tipo quindi 9,9,9): CASP = CASP-1$\cup$CASP-2$\cup$CASP-3, tale che per esempio: Pr(CASP-1|CASP) = $|$CASP-1$|$/$|$CASP$|$ = 9/27 = 1/3, *etc. etc.*?
Al termine di questa lunga Nota, ripetiamo, si tratta in ultima analisi unicamente di semplici calcoli aritmetici, e il metodo proposto per effettuarli, introducendo cioè il concetto di "albero" e i precedenti "disegnini", è soltanto un ausilio per l'intelletto, che si ritiene (auspica) possa essere conveniente.

Naturalmente, dopo che la scelta S1 è stata effettuata (e non solo contemplata come possibile tra le tre distinte S1, S2, S3), potremo trasformare il grafo [52] nel seguente, dove Pr(S1) è diventata 1, i rami con probabilità 0 sono stati eliminati, e tutte le altre probabilità (totali) sono state adeguatamente moltiplicate per 3:



[53]

Se torniamo alla DF, dobbiamo ancora modificare le probabilità totali nel precedente grafo, perché adesso Pr(P3) è diventata 1, mentre Pr(P2) è diventata 0, con tutte le conseguenze del caso (tra le quali, la "scomparsa" dell'intera seconda sezione del grafo (ricordiamo che la presenza di un 1 in un dato nodo implica tutti 1 **prima** di quel nodo, mentre quella di uno 0 implica tutti 0 **dopo**), ottenendo in conclusione il seguente "piccolo" albero (addirittura "piccolissimo" se eliminassimo dal disegno l'intera seconda sezione del grafo):



[54]

Quasi inutile rilevare che, posta qui attenzione ai due cammini S1-P3-A1 e S1-P3-A2, scopriamo che la probabilità di vincere confermando la scelta S1 è uguale a 1/3 = Pr(S1-P3-A1), mentre quella di vincere modificandola da S1 a S2 è uguale a 2/3 = Pr(S1-P3-A2) (se non annoia l'ulteriore ripetizione, S3 non è possibile dopo che è stata aperta la porta P3).

In definitiva, abbiamo infine "ritrovato" ciò che avevamo già detto in principio, ma in forma forse più ... elegante (e pure convincente, almeno si spera). Un lungo cammino per arrivare ad un risultato che in sostanza si poteva determinare parecchio prima facendo ricorso a qualche azzeccato ragionamento. Ciò nondimeno, la sua conferma per altra (algoritmica) via può essere a nostro parere ritenuta interessante, pure tenendo conto della circostanza che la proposta passeggiata ... tra gli alberi ha (forse) permesso di imparare (o di comprendere meglio) qualcosa e, come si dice, la soddisfazione consiste più nel viaggio che non nel raggiungimento della mèta...

UB, aprile 2024

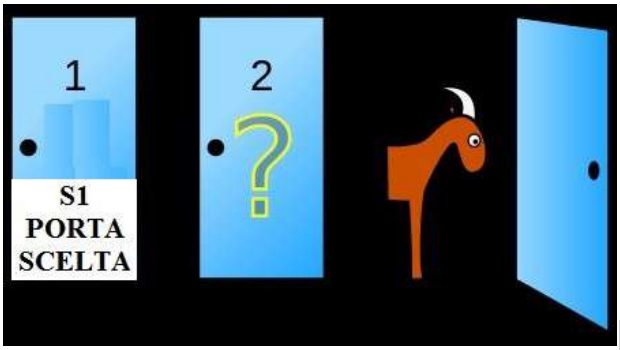
Breve Coda (maggio 2024) - Come già detto, abbiamo di proposito rinunciato ad entrare nel campo dei commenti al famoso paradosso, limitandoci a mostrare come esso si possa facilmente "risolvere" utilizzando semplici metodi di Calcolo delle Probabilità. Del resto, di commenti se ne trovano tanti in rete, e spesso anche interessanti (almeno in parte), per esempio:

https://chance.amstat.org/2022/11/monty-hall/
Monty Hall and the 'Leibniz Illusion'

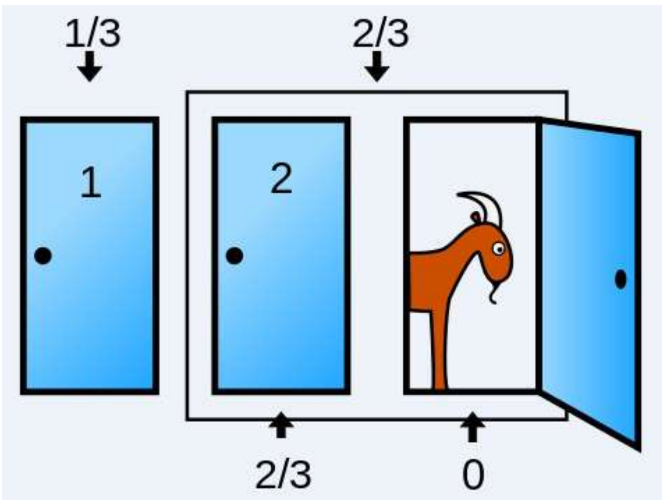
https://ima.org.uk/4552/dont-switch-mathematicians-answer-monty-hall-problem-wrong/
Don't Switch! Why Mathematicians' Answer to the Monty Hall Problem is Wrong

https://arxiv.org/pdf/1002.0651
The Monty Hall Problem is not a Probability Puzzle (It's a challenge in mathematical modelling)

A seguito di qualche scambio di idee susseguente alla prima pubblicazione in rete del presente scritto, ci pare però opportuno aggiungere un nostro breve commento all'unico "teorema" che vi si trova, prendendo spunto dalle due immagini che Wikipedia riporta per illustrarne la validità (immagini da noi leggermente ritoccate, e di significato che dovrebbe essere ormai del tutto evidente):



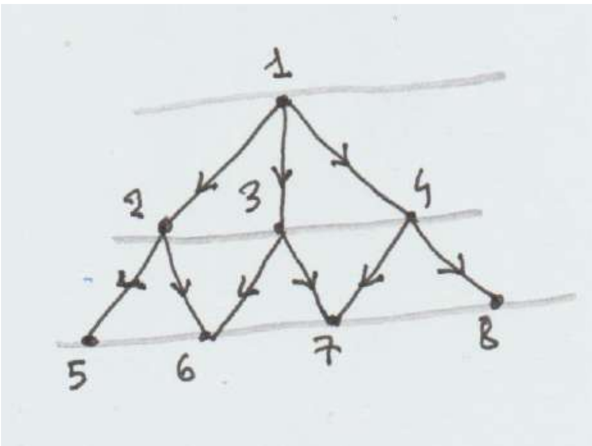
[55]



[56]

La domanda che ci poniamo è la seguente: il passaggio dall'immagine [55], quella **senza** le probabilità, alla [56], quella **con** le probabilità, è davvero univoco? La banale risposta è NO, se non si conosce come si è arrivati alla situazione descritta da [55]. Se P3 è stata aperta da un conduttore onnisciente, il quale per andare avanti a tutti i costi con lo spettacolo osserva rigorosamente le regole 1 e 2, allora le probabilità sono davvero quelle indicate in [56]. Se invece P3 è stata aperta **a caso** dal concorrente, diciamo per esempio dietro invito esplicito del conduttore, allora Pr(P3) è sì diventata 0, ma Pr(P1) e Pr(P2) sono diventate 1/2. In altre parole, non basta guardare [55] - come potrebbe fare qualsiasi interlocutore sopraggiunto soltanto a questo punto del discorso - per concludere con la validità di [56], bisogna conoscere la sequenza di eventi che a [55] ha condotto. Questa è invero una **sola** immagine, che può essere chiamata però a descrivere situazioni assai **diverse** tra loro, quindi con probabilità assai diverse, e decisioni razionali assai diverse da assumere da parte del concorrente.

Quanto dianzi osservato (banale) ci conduce anche all'osservazione generale che non è agli **stati** che bisogna guardare in certi casi, ma ai **cammini** che a quegli stati hanno portato. Nel modello ad alberi che abbiamo utilizzato, due cammini massimali diversi hanno sicuramente elementi finali diversi, i quali possono però corrispondere ad un medesimo "stato" (una sorta di "equivalenza" tra i nodi del grafo). Ecco così che appaiono forse più evidenti le ragioni per cui abbiamo proposto una siffatta modellizzazione: gli alberi sono in particolare dei grafi privi di **cicli**, ogni nodo diverso dalla radice dell'albero ammette un unico antecedente, non ci possono essere due nodi distinti ad un dato livello che sono collegati allo stesso nodo nel livello successivo, *etc. etc.*. Insomma, il seguente grafo **non** fa parte della famiglia degli alberi:



[57]

sebbene pur esso possa essere utilmente introdotto al fine per esempio di descrivere la possibile evoluzione di una certa partita a scacchi, ovvero la successione di diversi possibili stati di una scacchiera. Lo stato 3 può diventare lo stato 7 dopo una certa mossa, ma anche lo stato 4 può diventare lo stato 7 - la medesima configurazione dei pezzi sulla scacchiera - dopo un'altra data mossa. Rimanendo in tema di scacchi, ricordiamo che nella Nota 2 si è parlato della descrizione del "lavoro dell'intelletto di uno scacchista", ma va forse sottolineato che con il modello ad alberi quella che può essere descritta è la sequenza delle mosse, e non delle relative configurazioni della scacchiera. In altre parole, la successione: e4 (bianco) - e5 (nero) - d4 (bianco) è in un tale modello cosa ben diversa dalla: d4 (bianco) - e5 (nero) - e4 (bianco), nonostante si ottenga la stessa configurazione dei pezzi sulla scacchiera al termine delle 3 mosse (la freccia verticale nera indica la seconda mossa del bianco):


[58]

La prima successione di mosse corrisponde a quello che si dice un "gambetto del centro" (che non sappiamo ancora se verrà rifiutato oppure no), mentre la seconda ad un ... "gambetto Englund rifiutato", ossia nell'immagine che segue ad un bianco che, dopo la mossa e5 da parte del nero, risponde con e4, in luogo della più naturale pedone in d mangia pedone in e - un'apertura come si dice "eccentrica", con un seguito ancora più eccentrico, ma tant'è:


[59]