

## RETI PROBABILISTICHE E INTELLIGENZA ARTIFICIALE

**Carlo Berzuini**

*Dipartimento di Informatica e Sistemistica, Università di Pavia.*

*Si discute il ruolo della teoria della probabilità nei modelli computazionali del ragionamento umano e nella rappresentazione della conoscenza nei "sistemi esperti". Il punto di partenza è costituito dalla possibilità di rappresentare informazione probabilistica mediante modelli grafici, o reti, opportunamente quantificate. I due aspetti di un modello probabilistico, cioè la struttura delle indipendenze condizionali fra le variabili, da un lato, e la specifica delle distribuzioni probabilistiche in gioco, dall'altro, sono rappresentati separatamente. Il primo dei due aspetti, di natura qualitativa, viene espresso da un grafo orientato aciclico, mentre il secondo viene espresso sotto forma di probabilità condizionali associate ai nodi della rete. Ciò consente di svolgere elaborazioni parallele e distribuite, e di rappresentare in modo naturale processi di ragionamento deduttivo e abduttivo che coinvolgono variabili aleatorie. Questo lavoro presenta una rassegna dei recenti risultati di ricerche riguardanti (i) l'identificazione di indipendenze condizionali da una struttura grafica, (ii) i limiti di espressività di un modello grafico, e (iii) algoritmi per la propagazione dell'incertezza su strutture grafiche.*

### 1. INTRODUZIONE

L'idea di usare rappresentazioni grafiche per codificare informazione probabilistica risale forse al genetista S. Wright, il quale sviluppò il metodo chiamato *path analysis*. Nel 1934 egli pubblicò un lavoro su questo argomento, ma non ottenne successo presso gli statistici.

Quell'era era dominata da una visione esclusivamente quantitativa dell'analisi statistica e l'attesa di una nuova prospettiva durò fino agli anni '60. In questo periodo, infatti, rappresentazioni grafiche divennero un veicolo importante per esprimere modelli causali in economia (Wold, 1964), sociologia (Blalock, 1971) e psicologia (Duncan, 1975).

Più o meno nello stesso periodo, statistici come Goodman (1970) e Haberman (1974) si resero conto che alcune proprietà di decomposizione delle tabelle di contingenza, e una certa sottoclasse di modelli log-lineari di tali tabelle, si possono esprimere in modo ottimale attraverso *grafi*. Ciò segnò l'inizio di una graduale scoperta delle proprietà dei grafi come linguaggio per esprimere relazioni fra variabili. Se ne accorsero ad esempio i ricercatori impegnati nella teoria delle basi

di dati (Beeri et al, 1983). Infine i *diagrammi di influenza* (Howard e Matheson, 1981) (Shachter, 1988) (Smith, 1987) costituiscono un'altra componente di questa tradizione, e sono estremamente diffusi in ricerca operativa. Sviluppati per l'analisi decisionale essi distinguono fra tipi di nodi, fra cui i nodi decisionali e i nodi casuali.

Si è inoltre sviluppata una teoria, detta dei *campi Markoviani* (Markov Random Fields), che consente di codificare la struttura delle dipendenze insita in un modello statistico mediante una rete. Lauritzen (1982), ad esempio, ha applicato i campi Markoviani all'analisi di tabelle di contingenza mediante modello log-lineare. Altri usano campi Markoviani per rappresentare una struttura geometrica di adiacenze. Clayton (1988), ad esempio usa un campo Markoviano per rappresentare associazione spaziale fra rischi di malattia relativi a diversi distretti geografici. Egli usa il campo Markoviano come distribuzione a priori sul vettore dei rischi per un'analisi bayesiana delle variazioni di mortalità sul territorio. Geman & Geman (1984) avevano precedentemente applicato un concetto analogo all'elaborazione delle immagini.

Vicini a questa linea di pensiero, alcuni statistici, (vedi ad esempio Smith, 1988) usano il linguaggio dei grafi per analizzare modelli statistici complessi, e altri (Gelfand et al, 1988) sfruttano algoritmi basati su propagazione di probabilità su grafo per risolvere problemi computazionali nella statistica bayesiana (es. nei modelli gerarchici). L'utilità di tale approccio è particolarmente evidente nelle applicazioni in cui il modello coinvolge processi temporali e/o spaziali.

Infine, e qui entriamo nell'argomento più propriamente specifico di questa relazione, le reti probabilistiche sono state studiate in quanto modelli computazionali per il *ragionamento automatico* (Duda et al, 1976) (Pearl, 1986). In questo tipo di applicazione, la rete probabilistica rappresenta un sistema di interesse, ad esempio un'insieme di variabili relative ad un paziente da curare, o relative ad un sottosistema economico da controllare o su cui fare delle previsioni. L'obiettivo è quello di "ragionare" su tale sistema in base a informazioni e conoscenza imprecise (fare la diagnosi, scegliere la terapia, prevedere lo sviluppo futuro, ecc.). In questo tipo di applicazione apparve subito cruciale sviluppare algoritmi computazionalmente trattabili che consentissero di *propagare probabilità* su una rete. Mediante propagazione, l'impatto di informazione su alcuni nodi della rete si trasmette ad altri nodi che rappresentano *ipotesi* di interesse, con il risultato di mantenere il grado di credibilità di tale ipotesi aggiornato in funzione dell'evidenza che si rende via via disponibile.

Ora, proprio grazie al fatto che le reti probabilistiche collocano l'informazione necessaria a compiere le inferenze accanto alle variabili coinvolte nell'inferenza,

gli algoritmi di propagazione possono essere implementati in modo *distribuito*, e quindi compatibile con le esigenze computazionali dei moderni sistemi intelligenti. Oltre alla possibilità di elaborazione locale, le reti hanno il vantaggio della *modularità*. Questa consiste nel poter formulare un modello incrementalmente e gradualmente, aggiungendo informazione “qui e là”, “oggi e domani” nella rete che lo rappresenta. Ciò rende le reti probabilistiche adatte in ambienti *dinamici*, in cui cioè nuovi pezzi devono venire aggiunti, e quelli vecchi modificati via via che si rende disponibile nuova conoscenza. L'applicazione della teoria della probabilità ai “sistemi esperti” costituisce appunto l'argomento del presente lavoro. Esso infatti presenta una rassegna dei recenti risultati di ricerche riguardanti:

- (i) l'identificazione di indipendenze condizionali da una struttura grafica,
- (ii) i limiti di espressività di un modello grafico,
- (iii) algoritmi per la propagazione dell'incertezza su strutture grafiche.

## 2. GRAFI ORIENTATI ACICLICI E STRUTTURE DI INDIPENDENZA CONDIZIONALE

Un sistema probabilistico intelligente deve rispecchiare le rappresentazioni mentali usate dall'uomo nei ragionamenti. Esso deve incorporare modelli del mondo reale che riflettono esplicitamente e vividamente i cammini mentali percorsi da un'esperto umano nella risoluzione del problema di interesse. Ad esempio: un programma di aiuto alla diagnosi clinica che si limita ad ordinare ipotesi diagnostiche applicando ai sintomi di ingresso la formula di Bayes dell'indipendenza *non* risponde secondo noi al requisito sopra menzionato. Quand'anche azzeccasse ogni diagnosi, rimarrebbe sostanzialmente incapace di rappresentare e manipolare le relazioni qualitative che troviamo in una normale spiegazione diagnostica. Questa, infatti, contiene tipicamente una molteplicità di concetti e informazioni: sintomi, diagnosi, ipotesi intermedie, stati funzionali, astrazioni qualitative, e così via.

Un sistema intelligente deve normalmente gestire una complessa rete di relazioni fra tanti *pezzi di informazione*. In una qualsiasi attività di ragionamento, ogni passo del processo deduttivo (i logici distinguono fra deduzione, abduzione e induzione) richiede in linea di principio la presa in esame dell'intero insieme delle informazioni contenute nella base di conoscenza. La miriade delle possibili combinazioni sfuggirebbe al controllo se il sistema non fosse capace di identificare, in ogni fase del ragionamento, quali, fra i fatti contenuti nella base di conoscenza, sono *irrilevanti* nel dato contesto. L'articolazione del ragionamento è possibile solo se, in ognuno dei contesti che il ragionamento stesso crea dinamicamente, l'insieme

delle informazioni di cui si deve tener conto viene ridotto scartando quelle irrilevanti.

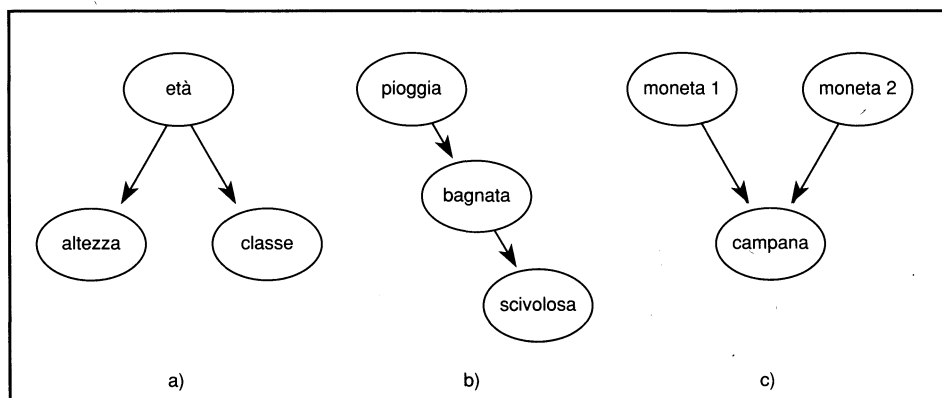
Nella teoria della probabilità, la nozione di rilevanza si identifica con il concetto di *indipendenza condizionale*. Nel seguito useremo  $i(X, Z, Y)$  per indicare indipendenza fra  $X$  e  $Y$  data  $Z$ . Date tre variabili aleatorie  $X, Y$  e  $Z$ , si definisce:

$$i(X, Z, Y) \text{ se e solo se } P(x/y, z) = P(x/z) \quad (1)$$

per tutti i valori  $y$  e  $z$  tali che  $P(y, z) > 0$ . La nozione di indipendenza condizionale si può interpretare più intuitivamente in termini di *rilevanza informazionale*. Infatti, l'affermazione " $X$  e  $Y$  sono condizionalmente indipendenti data  $Z$ " significa "una volta nota  $Z$ , il valore di  $Y$  (risp.  $X$ ) è del tutto *irrilevante* per predire il valore di  $X$  (risp.  $Y$ )".

Un insieme di relazioni di indipendenza condizionale si può rappresentare mediante un *Grafo Orientato Aciclico* (GOA). Se i nodi di un GOA rappresentano variabili aleatorie e i lati (freccie) rappresentano influenze causali dirette, allora la topologia del grafo corrisponde a una lista di asserzioni di indipendenza condizionale circa quelle variabili.

Consideriamo ad esempio i GOA elementari riportati nelle figure 1a–c. La Fig. 1a implica che *altezza* e *classe scolastica* di un bambino sono *marginalmente* dipendenti, ma *condizionalmente* indipendenti data l'*età* del bambino. Questa interpretazione deriva dall'attribuire ai lati del GOA un significato di influenza *causale*: *altezza* e *classe* sono "causate" da un fattore comune, l'*età*, che le rende dipendenti "nella popolazione"; tuttavia, una volta appresa l'*età* di un determinato bambino, la sua *altezza* e la sua *classe* diventano mutuamente irrilevanti, poiché



**Fig. 1:** Queste strutture causali hanno una semantica basata sul concetto di indipendenza condizionale: *altezza* e *classe* sono marginalmente dipendenti ma condizionalmente indipendenti data *età*. Anche *pioggia* e *scivolosa* sono marginalmente dipendenti ma condizionalmente indipendenti data *bagnata*. *Moneta 1* e *moneta 2* sono marginalmente indipendenti ma condizionalmente dipendenti data *campana*.

età riassume tutto ciò di cui, il modello riportato, devo essere a conoscenza per predire *altezza* e *classe*. Il termine “causale” che stiamo usando richiederebbe una definizione “epistemologica” formale, ma non entriamo in tale questione.

La Fig. 1b contiene tre variabili dicotomiche, che indicano se è piovuto (*pioggia*), se la strada è bagnata (*bagnata*) e se la strada è scivolosa (*scivolosa*). Secondo il diagramma, la variabile *bagnata* “media” completamente l'effetto causale della variabile *pioggia* sulla variabile *scivolosa*. Ne segue, intuitivamente, che il sapere che la strada è *bagnata* rende l'informazione circa la *pioggia* irrilevante per dedurre se la strada è *scivolosa*, ovvero vale  $i(\text{pioggia}, \text{bagnata}, \text{scivolosa})$ .

Nella Fig. 1c *Moneta 1* e *moneta 2* rappresentano gli esiti dei lanci di due monete (ben bilanciate). Immaginiamo una campana che suona ogni qualvolta gli esiti sono identici, e definiamo la variabile dicotomica *campana*, che indica appunto se la campana suona o meno. *Moneta 1* e *moneta 2* rappresentano eventi normalmente indipendenti, ma diventano mutuamente dipendenti non appena apprendiamo il valore di *campana*. Infatti, se ignoriamo la campana gli esiti dei lanci sono mutuamente indipendenti, ma se la notiamo allora gli esiti interagiscono, poiché in tal caso l'apprendere l'esito di un lancio dovrebbe modificare la nostra opinione circa l'altro.

L'ultimo esempio è particolarmente significativo, poiché mostra che un GOA non rappresenta solo dipendenze “genuine” ma anche dipendenze *indotte* da ipotetiche osservazioni.

È evidente a questo punto il bisogno di una regola generale per estrarre asserzioni di indipendenza da un grafo orientato. Un criterio grafico estremamente importante, chiamato *d-separazione*, è il seguente:

*Definizione (Pearl, 1988):* Se  $X$ ,  $Y$ , e  $Z$  sono tre sottoinsiemi disgiunti di nodi in un GOA, si dice che  $Z$  *d-separa*  $X$  da  $Y$ , in breve  $i(X, Z, Y)_D$ , se non vi è alcun cammino fra un nodo in  $X$  e un nodo in  $Y$  lungo il quale (1) ogni nodo su cui convergono due frecce è o ha un successore in  $Z$ , e (2) ogni altro nodo non appartiene a  $Z$ . Un cammino che soddisfa le due condizioni appena menzionate si dice “attivo”, altrimenti si dice “bloccato”.

In Fig. 2, ad esempio, fra  $B$  e  $D$  vi sono due cammini, cioè  $B \leftarrow A \rightarrow D$  e  $B \rightarrow E \rightarrow F \leftarrow D$ . L'insieme  $\{A, F\}$  blocca  $B \leftarrow A \rightarrow D$ , ma attiva  $B \rightarrow E \rightarrow F \leftarrow D$ . Perciò  $B$  and  $D$  non sono *d-separati* da  $\{A, F\}$ . Se invece poniamo  $Z = \{A, E\}$ ,  $Z$  non attiva alcun cammino tra  $B$  e  $D$ , perciò in tal caso  $Z$  *d-separa*  $B$  e  $D$ .

Se ora ci riferiamo agli esempi di GOA in Fig. 1, possiamo verificare che le indipendenze condizionali che avevamo identificato in tali grafi corrispondono esattamente alle *d-separazioni*. Più precisamente, ogni *d-separazione*  $i(X, Z, Y)_D$

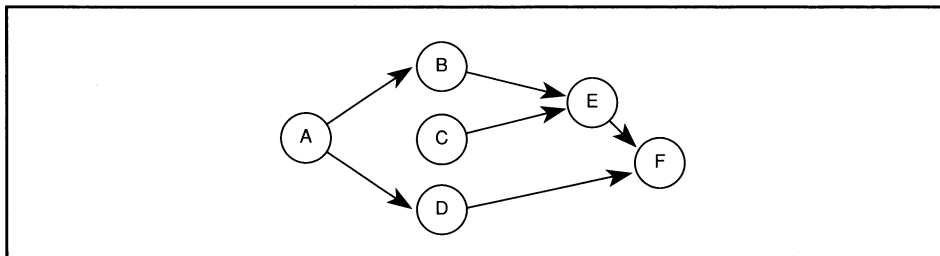


Fig. 2: Semplice esempio di GOA.

rappresentata in un grafo corrisponde a una indipendenza condizionale  $i(X, Z, Y)$ . Di conseguenza, se  $Z$  rappresenta il sottinsieme delle variabili *note* nel generico passo del ragionamento, allora tutte le coppie di nodi  $(X, Y)$  per cui vale  $i(X, Z, Y)_D$  rappresentano entità mutuamente *irrilevanti* in quel punto del ragionamento. Ad esempio, secondo la Fig. 1(b), quando non sappiamo se la strada è *bagnata*, *pioggia* e *scivolosa* non sono  $d$ -separate (il cammino tra di esse è “attivo”), il che indica che esse sono mutuamente rilevanti. Tuttavia, non appena apprendiamo, ad esempio, che la strada è bagnata, allora in virtù di  $i(\text{pioggia}, \text{bagnata}, \text{scivolosa})_D$ , deduciamo che, data quell’informazione, non ho bisogno di chieder se è piovuto per dedurre che la strada è scivolosa (o viceversa).

Per concludere, in questo paragrafo abbiamo visto come i GOA si prestino alla descrizione di conoscenza causale e come le relazioni causali rappresentate in un GOA si possano interpretare coerentemente in termini di indipendenza condizionale. Grazie a ciò, e grazie all’esistenza di un criterio grafico come la  $d$ -separazione, i GOA acquistano, al di là di una funzione puramente mnemonica, anche una funzione di supporto ad algoritmi che identificano fra le variabili rappresentate quelle rilevanti per un certo passo di ragionamento. Questa funzione, come vedremo, è molto utile nei compiti di inferenza e di spiegazione eseguiti da un sistema esperto.

### 3. I LIMITI DI ESPRESSIVITÀ DEI GRAFI ORIENTATI ACICLICI

I semplici esempi di GOA riportati nel paragrafo precedente sono stati analizzati come strutture esclusivamente *qualitative*, prive di specificazioni distribuzionali, o numeriche. Tuttavia essi possono essere *quantificati*, specificando la distribuzione congiunta delle variabili in essi inclusi. Una volta quantificato, un GOA diventa un vero e proprio modello statistico, espresso sotto forma di *rete probabilistica*.

Quantificare un GOA è concettualmente semplice: basta specificare per ogni nodo  $X$ , a cui giungono frecce che emanano da un insieme  $\Pi_X$  di nodi “genitori”, le

probabilità condizionali  $P(x / \Pi_x)$  per ogni valore  $x$  di  $X$  e ogni possibile assegnazione di valori a  $\Pi_x$ .

A titolo illustrativo consideriamo il GOA riportato in Fig. 3 (si tratta di un esempio con funzione puramente illustrativa, da non prendersi quindi troppo sul serio da un punto di vista clinico). I nodi del grafo esprimono eventi che possono accadere ad un paziente in conseguenza di una *metastasi* dovuta a una precedente storia di tumore. La *metastasi* può causare lo sviluppo di un secondo tumore, al cervello, e/ o *ipercalcemia*, e sia *ipercalcemia* che *tumore-cerebrale* possono, indipendentemente, causare *coma*. Infine, il *tumore-cerebrale* può provocare *forti-emicranie*.

Ogni nodo è quantificato, come dicevamo, da probabilità condizionali dati i nodi da cui emanano frecce verso di esso. Nella figura viene riportata, a titolo di esempio, la tabella delle probabilità condizionali associata al nodo *coma*. Esse comunicano, all'incirca, che la conferma di uno qualsiasi, o di entrambi, gli stati *ipercalcemia* e *tumore cerebrale* rende *coma* "molto probabile", mentre la negazione di entrambi lo rende "estremamente poco probabile".

Una specificazione di probabilità secondo la regola appena vista è completa e consistente perché il prodotto delle probabilità condizionali sull'insieme dei nodi:

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^N P(X_i | \Pi_{X_i}) \quad (2)$$

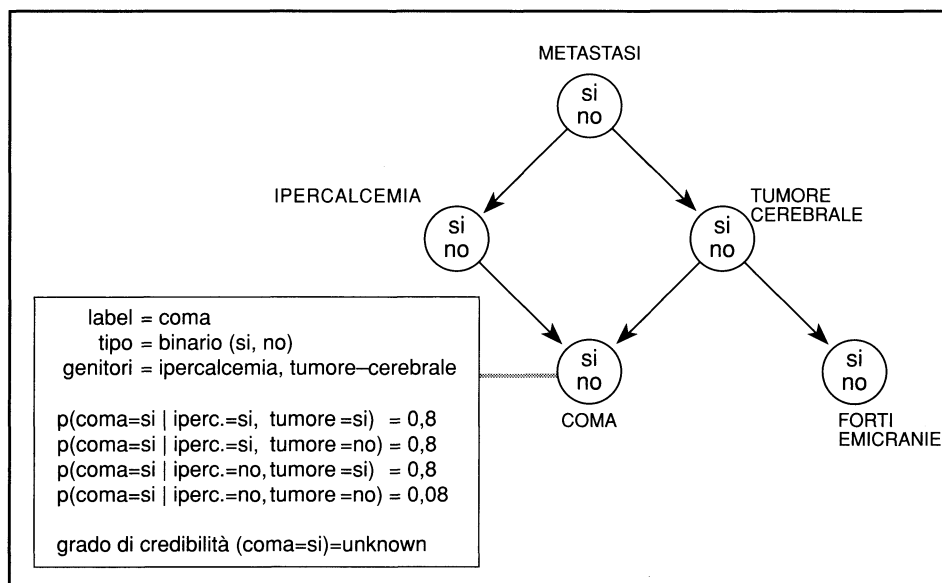


Fig. 3: Semplice esempio di Rete Causale Probabilistica (RCP).

costituisce (modulo una costante) la densità congiunta delle variabili rappresentate. Ed è importante osservare che la densità congiunta ottenuta in questo modo è consistente con le indipendenze condizionali estratte dalla struttura qualitativa del GOA sottostante mediante il criterio della  $d$ -separazione. Nel grafo in figura 3 le frecce rappresentano le influenze "causali" dirette che noi percepiamo fra le variabili in gioco. Va sottolineato il termine "dirette": ad esempio, *metastasi* influenza causalmente *coma*, ma indirettamente, per cui tra i due nodi è assente la freccia. Tale assenza equivale ad asserire che i due nodi sono condizionalmente indipendenti date *ipercalcemia* e *tumore cerebrale*. Le frecce assenti rappresentano dunque indipendenze condizionali, e la (2) è perfettamente consistente con tali indipendenze. Il criterio della  $d$ -separazione si può a questo punto interpretare come uno strumento per estrarre dal grafo tutte e sole le indipendenze che seguono logicamente da quelle implicite nelle frecce mancanti.

A questo punto è opportuno ribaltare la prospettiva rispetto al paragrafo precedente. Invece di partire da un grafo che riflette la nostra percezione delle relazioni causali in gioco e successivamente quantificarlo in accordo con la (2), partiamo da un modello probabilistico  $M$ , definito da una distribuzione congiunta sulle variabili in gioco. Ci chiediamo poi: possiamo codificare la struttura delle indipendenze condizionali imposta da  $M$  mediante un GOA  $D$  in cui i nodi corrispondono alle variabili incluse in  $M$ , nel senso che l'applicazione del criterio della  $d$ -separazione porta a tutte e sole le indipendenze implicate da  $M$ ? Questo quesito si può formalizzare introducendo in concetto di GOA – isomorfismo:

*Definizione (Pearl, 1988): Un modello probabilistico  $M$  si dice GOA-isomorfo se esiste un GOA  $D$  i cui nodi corrispondono 1 a 1 alle variabili in  $M$  e che è una trasformazione perfetta di  $M$  relativamente alla  $d$ -separazione, cioè tale per cui se  $X$ ,  $Y$ , e  $Z$  sono tre sottoinsiemi disgiunti di nodi in  $D$ , allora:*

$$i(X, Z, Y)_D \Leftrightarrow i(X, Z, Y) \quad (3)$$

Il quesito che ci eravamo posti sopra si può dunque porre nei termini seguenti: sotto quali condizioni un dato modello probabilistico è GOA-isomorfo?

Ancora una volta la risposta è stata data da Pearl (1988), che ha definito 7 assiomi che devono essere soddisfatti da un modello probabilistico per essere GOA-isomorfo. Si tratta di condizioni necessarie, ma non sufficienti (forse il numero delle condizioni sufficienti è infinito). Una di queste condizioni è la transitività debole (le lettere greche significano nodi singoli, le altre indicano insiemi di nodi):

$$i(X, Z, Y) \& i(X, Z \cup \gamma, Y) \Rightarrow i(X, Z, \gamma) \text{ or } i(\gamma, Z, Y) \quad (4)$$



la quale significa che se  $X$  e  $Y$  sono sia marginalmente che condizionalmente indipendenti data  $\gamma$ , è impossibile che sia  $X$  che  $Y$  siano indipendenti da  $\gamma$ . Un'ulteriore condizione necessaria è la cordalità:

$$i(\alpha, \gamma \cup \delta, \beta) \& i(\gamma, \alpha \cup \beta, \delta) \Rightarrow i(\alpha, \gamma, \beta) \text{ or } i(\alpha, \delta, \beta) \quad (5)$$

Diamo ora un esempio di modello non GOA-isomorfo. Consideriamo il seguente brano:

*Una mattina, Maria telefona al suo ragazzo, Giovanni, pregandolo di raggiungerla a casa. Non appena ricevuta la chiamata, Giovanni prende la macchina e parte, diretto verso la casa di Maria. Il tempo richiesto per arrivare aumenta se l'usuale traffico mattutino ha creato una coda sotto l'abitazione di Giovanni. Se Giovanni tarda, Maria si preoccupa, e gli telefona una seconda volta.*

Il pezzo di “conoscenza” sopra riportato si potrebbe modellizzare introducendo variabili aleatorie continue, chiamate “date”, che rappresentano gli istanti di tempo in cui avvengono gli eventi di interesse, cioè *prima chiamata* ( $T$ ), *parte* ( $P$ ), *arriva* ( $A$ ), *formazione della coda* ( $C$ ), *seconda chiamata* ( $S$ ).

Fra i valori (non-negativi) che tali date possono assumere includiamo il valore “infinito” ( $\infty$ ) per indicare “mai”. La struttura delle indipendenze condizionali fra le date del nostro esempio è piuttosto complicata. Limitiamoci al caso particolare in cui possiamo assumere  $S \neq \infty$  (cioè  $M$ . ha effettivamente compiuto una seconda chiamata). La distribuzione  $p(T, P, S, A \mid S \neq \infty)$  dovrebbe essere consistente con:

$$i(P, T \cup A, S) \& i(T, P \cup S, A) \quad (6)$$

dove il primo termine esprime il fatto intuitivo che, fissati gli istanti in cui  $M$ . chiama la prima volta e quello in cui  $G$ . arriva, allora l'istante in cui  $M$ . chiama una seconda volta non è influenzato dall'istante in cui  $G$ . è partito. Il secondo termine esprime il fatto intuitivo che, se sappiamo se/quando  $G$ . è partito e quando  $M$ . ha chiamato una seconda volta, allora per predire se/quando  $G$ . è arrivato non ci è utile sapere quando  $M$ . ha chiamato la prima volta. Se è vera la condizione di cordalità (5), dalla (6) deduciamo che almeno una delle seguenti asserzioni è vera:

$$i(P, T, S) \quad (7)$$

$$i(P, A, S) \quad (8)$$

Le (7)–(8) sono, però, contrarie alla nostra intuizione. La (7) è contraddetta dal fatto che, fissato l'istante  $T$  in cui  $M$ . chiama la prima volta, il variare di  $P$  (cioè dell'istante in cui  $G$ . parte) influenza la distribuzione di probabilità di  $A$  (cioè

dell'istante in cui  $G$ . arriva) e quindi indirettamente quella di  $S$  (cioè dell'istante in cui  $M$ . chiama una seconda volta), essendo  $S < A$ . La (8) è contraddetta da considerazioni analoghe. Concludiamo dunque che le indipendenze condizionali fra le nostre date violano la condizione di cordalità e quindi, per quanto detto precedentemente, non possiamo rappresentare tutte simultaneamente in un GOA, a meno di non introdurre nel GOA indipendenze indebite. Il lettore potrebbe provare a formare una serie di GOA con nodi  $(T, P, A, S)$  (tralasciando  $C$  per semplicità), per verificare, armato del criterio della  $d$ -separazione, che la soddisfazione della (6) e la negazione delle (7)–(8) non sono mai ottenibili simultaneamente.

Come possiamo superare questa deficienza di espressività? L'unica soluzione (Berzuini, 1989) sta nell'*inventare* nuove variabili da inserire opportunamente nella rete per portare alla luce tutte e sole le indipendenze che esistono tra le variabili rappresentate. Nel nostro esempio possiamo ricorrere a una *data potenziale* che rappresenta l'istante in cui  $M$ . farebbe la seconda chiamata *supposto che*  $G$ . non arrivi mai. Questa data permette di strutturare il GOA che descrive il nostro esempio come in figura 4, dove l'istante "potenziale" della seconda chiamata è denotata con simbolo  $X$ . La maggior parte dei nodi si può quantificare tramite funzioni di "rischio istantaneo" analoghe a quelle usate negli studi di sopravvivenza; ad esempio ad  $X$  possiamo associare una funzione di rischio  $h(X | T)$  che esprime l'evoluzione temporale della propensione di  $M$ . a compiere la seconda chiamata su una scala temporale che ha come origine proprio l'istante  $T$  della prima chiamata. L'istante *effettivo* della seconda chiamata ( $S$ ) sarà uguale a  $\infty$  se  $X > A$ , mentre sarà uguale a  $X$  se  $X < A$ . Si può verificare come su tale grafo le indipendenze in (6), *ma non* la (7) e la (8), corrispondano a effettive  $d$ -separazioni.

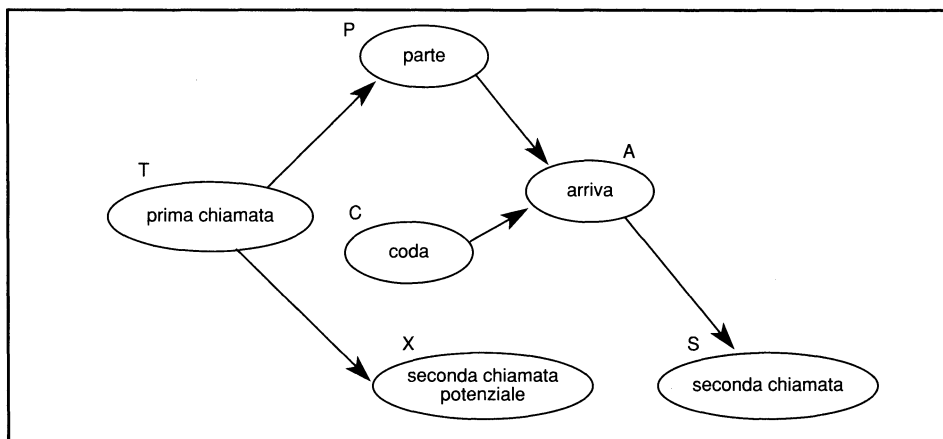


Fig. 4: L'esempio della "coda".

#### 4. ESEMPIO

Utilizzando il criterio della  $d^*$ -separazione, programma su calcolatore può usare un GOA per stabilire, qualsiasi sia lo stato corrente della conoscenza, quali informazione sono rilevanti in merito ad una qualunque inferenza si voglia eseguire. Ora si immaginino un programma e un utente che dialogano circa il nostro "esempio della coda" per dedurre da evidenza disponibile se Giovanni ha incontrato o meno coda. Mentre l'utente usa il suo buon senso, il programma potrebbe usare il criterio della  $d^*$ -separazione per derivare dal grafo riportato in figura 4 asserzioni compatibili con il buon-senso umano. Presentiamo nel seguito un esempio di come potrebbe apparire questa conversazione, racchiudendo le asserzioni di  $d^*$ -separazione fra parentesi quadre.

UTENTE: Posso dedurre se G. ha trovato la *coda* se so che M. ha fatto la *prima chiamata* a G. alle 8.00, e che G. è *partito* alle 8.06?

PROGRAMMA: No, a meno che tu non sappia anche quando G. è *arrivato* [la variabile "coda" rimane  $d^*$ -separata dalle variabili "prima chiamata" e "parte" fino a che non si conosce il valore della variabile "arriva"].

UTENTE: In realtà, non ho questa informazione.

PROGRAMMA: Allora potrebbe essere utile sapere se/quando M. ha fatto la sua *seconda chiamata* [il cammino fra le variabili *seconda chiamata* – *arriva* – *coda* è  $d^*$ -attivo fintantoché la variabile "arriva" è incognita].

UTENTE: Io so  $X=S=8.28$  e  $T=8.00$ , cioè che M. ha fatto la *seconda chiamata* 28 minuti dopo la prima, perché non vedeva G. arrivare. Secondo me, ciò significa che il viaggio di G. è durato un tempo particolarmente alto. Ciò mi suggerisce che G. ha trovato *coda*. Ho tenuto conto di tutta l'informazione rilevante?

PROGRAMMA: No, trascuri evidenza circa l'istante in cui G. è *partito*. [ $\{X, S, T\}$  non  $d^*$ -separano "coda" da "parte"].

UTENTE: Sì, è vero, ho dimenticato di tener conto che G. è partito alle 8.06. Da questa informazione, e dal sapere che  $S=8.28$ , deduco che il viaggio è durato almeno 22 minuti. Comunque un tempo alto. Di conseguenza, riconfermo la mia precedente opinione circa la presenza di coda nel momento in cui G. è partito.

## 5. IL RAGIONAMENTO PROBABILISTICO

Nella rimanente parte di questo lavoro esaminiamo vari aspetti dell'inferenza probabilistica su reti. Per "inferenza probabilistica" intendiamo qui un processo che inizia con l'assegnazione di determinati valori ai nodi della rete che rappresentano proposizioni o variabili osservate, che costituiscono l'*evidenza*, per arrivare a determinare le *probabilità a posteriori* associate ai nodi della rete che corrispondono a proposizioni/variabili che costituiscono le *ipotesi* di interesse. Tali probabilità si indicano con la notazione:

$$p(\text{ipotesi} | \text{evidenza})$$

cioè "probabilità che sia vera l'ipotesi condizionatamente alla totalità dell'evidenza da cui sono partito". Tali probabilità a posteriori vengono anche chiamate *gradi di credibilità*.

Per scopo illustrativo, consideriamo la rete in fig. 5, la quale indica che lo scompenso cardiaco può produrre edema malleolare, e che la nefrosi può produrre sia edema malleolare che proteinuria (l'esempio è volutamente semplicistico per semplificare l'esposizione). Sono anche riportate in figura probabilità condizionali che racchiudono "informazione epidemiologica" relativa alle variabili in gioco. La seguente inferenza:

*"vidi subito che il paziente presentava edema malleolare, allora sospettai uno scompenso. Poi seppi che il paziente presentava proteinuria, e capii che non si trattava di quello: il paziente in realtà soffriva di nefrosi"*

è intuitiva se pensiamo al modello in Fig. 5. Infatti "scompenso" e "nefrosi" sono "spiegazioni" del sintomo "edema" fra loro *in competizione*. Di conseguenza, poiché l'evidenza di "proteinuria" rende pressoché certa l'ipotesi di "nefrosi", ne risulta da essa discredita l'ipotesi di "scompenso". Questo ragionamento, ovvio alla luce del nostro buon senso, può essere trascritto in termini probabilistici come segue:

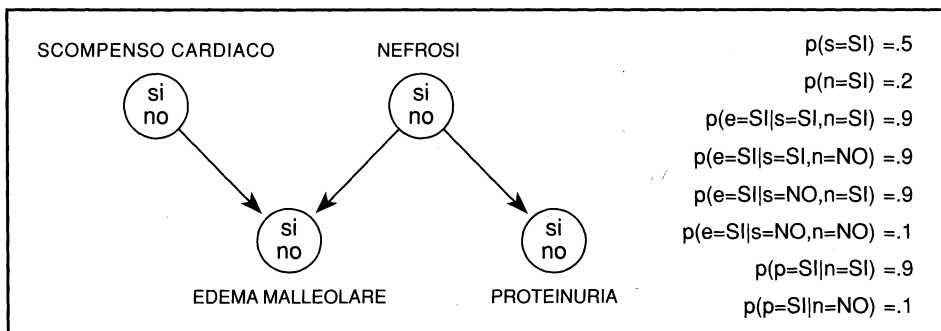


Fig. 5: Semplice esempio di Rete Causale Probabilistica a due soli livelli.

- (a) l'unica evidenza di cui dispongo inizialmente, cioè la presenza di edema, rende la probabilità a posteriori di "scompenso", cioè  $p(s=SI|e=SI)$  alta;
- (b) di conseguenza, relativamente a tutto ciò che so, cioè la presenza di un edema, è razionale che io creda nell'ipotesi "scompenso cardiaco";
- (c) ora dispongo di un'evidenza accresciuta, cioè informazione sull'edema e sulla proteinuria, che rende  $p(s=SI|e=SI, p=SI)$  bassa;
- (d) di conseguenza, relativamente a tutto ciò che so, cioè la presenza di edema e di proteinuria, non è razionale che io creda nell'ipotesi "scompenso cardiaco"

Può la teoria della probabilità render conto della dinamica del ragionamento sopra riportato? Cerchiamo di scoprirlo con un pò di calcoli.

Proviamo innanzitutto a verificare (a) sulla base dell'informazione (struttura della rete + probabilità condizionali) contenuta in Figura 5, utilizzando gli assiomi della teoria della probabilità:

$$p(s=SI|e=SI) = \alpha \{ p(+e|+s, +n)p(+s)p(+n) + p(+e|+s, -n)p(+s)p(-n) \} = \alpha 0,45 \quad (9)$$

$$p(s=NO|e=SI) = \alpha \{ p(+e|-s, +n)p(-s)p(+n) + p(+e|-s, -n)p(-s)p(-n) \} = \alpha 0,05 \quad (10)$$

dove la costante di normalizzazione è data da  $\alpha = (.45+.05)^{-1} = 2$ , da cui segue

$$p(s=SI|e=SI) = 0,9 \quad (11)$$

che traduce, in effetti, la proposizione (a).

Per verificare la proposizione (c), invece, partiamo dalla fattorizzazione della densità congiunta delle quattro variabili:

$$P(s, n, e, p) = p(s)p(n)p(e|s, n)p(p|n) \quad (12)$$

e otteniamo le seguenti probabilità a posteriori mediante marginalizzazione:

$$\begin{aligned} p(s=SI|e=SI, p=SI) &= \beta P(+s, +n, +e, +p) + P(+s, -n, +e, +p) = \\ &= \beta p(+s)p(+n)p(+e|+s, +n)p(+p|+n) + \beta p(+s)p(-n)p(+e|+s, -n)p(+p|-n) = \beta 0,117 \end{aligned} \quad (13)$$

$$\begin{aligned} p(s=NO|e=SI, p=SI) &= \beta P(-s, +n, +e, +p) + P(-s, -n, +e, +p) = \\ &= \beta p(-s)p(+n)p(+e|-s, +n)p(+p|+n) + \beta p(-s)p(-n)p(+e|-s, -n)p(+p|-n) \end{aligned} \quad (14)$$

da cui si ottiene  $\beta = (.117+.085)^{-1} = 4.95$ , e quindi infine, sostituendo il valore calcolato di  $\beta$  nella (13), si ottiene  $p(s=SI|e=SI, p=SI) = 0.57$ , cioè un grado di credibilità marcatamente diminuito rispetto a quello riportato nella (11). Quest'ultima conclusione traduce, in effetti, la proposizione (c).

Quello sopra è solo uno degli infiniti esempi che confermano come la teoria della probabilità non produca mai risultati contro-intuitivi. Ci si potrebbe dilungare

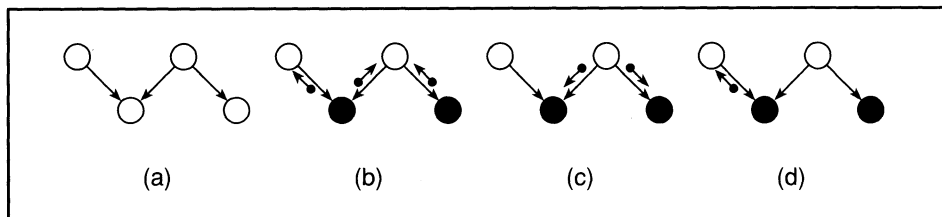
mostrando come i “patterns” qualitativi più comuni nel nostro ragionamento quotidiano siano incorporati in modo naturale nella sua sintassi. Fra questi “patterns” menzionano il ragionamento deduttivo, il ragionamento abduttivo e, soprattutto, la non-monotonicità. Quest’ultimo aspetto si riferisce alla possibilità che il processo di revisione dei gradi di credibilità innescato dall’arrivo di nuova evidenza porti alla “ritrattazione” di conclusioni precedentemente raggiunte, come avviene nell’esempio appena riportato. Questo aspetto non è previsto dalla logica classica, essenzialmente “monotonica”, dove l’insieme delle conclusioni deducibili cresce costantemente con l’aggiungersi di nuova informazione (nuovi “assiomi” non possono soprafare “teoremi” stabiliti in precedenza). Questa carenza si traduce in notevoli difficoltà incontrate nei sistemi esperti non-probabilistici. Forse, l’adozione dei “patterns” menzionati sopra come assiomi del nostro comportamento intelligente renderebbe praticamente inevitabile adottare la teoria della probabilità come formalismo di base per un modello computazionale di tale comportamento. L’indagine sulla compatibilità fra probabilità e intuizione umana ha costituito un tema di ricerca preferenziale per molti filosofi, fra cui (Ramsey, 1931), (De Finetti, 1937), e (Cox, 1946).

## 6. PROPAGAZIONE DELL’INCERTEZZA MEDIANTE ELABORAZIONI LOCALI

Le formule (9)–(14) usate per l’esempio di inferenza nel paragrafo precedente partono da una probabilità congiunta fattorizzata, e poi marginalizzano. Si tratta di calcoli banali, che però diventerebbero formidabili su una rete appena più grande. Immaginate di automatizzare un procedimento di questo tipo in modo tale da funzionare indipendentemente dal particolare tipo di rete; ciò significherebbe delegare ad un programma “supervisore” l’esecuzione di calcoli grandiosi qualunque sia il grado di complessità dell’inferenza richiesta.

Le indipendenze condizionali incorporate nella rete creano la possibilità di procedure ben più efficienti, che tengono conto della struttura per ottimizzare la quantità di calcoli. L’idea è quella di usare le reti probabilistiche come architetture computazionali per l’inferenza attraverso calcoli di tipo locale che avvengono tramite scambio di informazione fra i nodi.

L’impatto di ogni nuova evidenza su un qualche nodo della rete può essere visto come una perturbazione che si propaga attraverso la rete mediante scambio di messaggi fra nodi adiacenti. Come risultato di questa propagazione, il grado di credibilità (probabilità a posteriori) di tutti i nodi della rete “non colpiti” direttamente da evidenza viene modificato in ragione di questa, e consistentemente con gli assiomi della teoria della probabilità. Questo procedimento, chiamato “propagazio-



**Fig. 6: Rappresentazione schematica delle fasi di un processo di propagazione dell'incertezza sulla RCP già mostrata in figura 2.**

ne dell'incertezza", è frutto di elaborazioni locali compiute dai processori insiti nei nodi della rete, i quali comunicano tra di loro attraverso canali corrispondenti ai lati stessi della rete. Tutto ciò avviene, almeno in linea di principio, senza alcuna supervisione esterna, e attraverso procedure del tutto indipendenti dalla particolare istanza di rete.

Per esempio, immaginiamo che l'impatto dell'evidenza "*edema+proteinuria*" sulla nostra rete—esempio attivi il processo in quattro fasi descritto in Fig. 6. I "messaggi" da nodo a nodo sono disegnati come palline con freccia, che chiameremo "gettoni". Questi contengono il risultato di elaborazioni numeriche svolte autonomamente dai processori insiti nei nodi sui messaggi ricevuti, usando l'informazione contenuta nelle distribuzioni condizionali "locali" ad essi associate, ma non entreremo nel dettaglio di questi contenuti, poiché qui ci interessa solo comunicare un'idea della dinamica generale del processo di propagazione. La propagazione assomiglia a un gioco:

- (a) la rete è in "equilibrio";
- (b) l'evidenza colpisce i due nodi anneriti nel disegno. L'impatto attiva i processori insiti in questi due nodi, i quali reagiscono elaborando gettoni e spedendoli in direzione dei rispettivi "genitori";
- (c) i genitori, attivati dai gettoni ricevuti, si preoccupano a loro volta di elaborare nuovi gettoni e di spedirli verso nodi contigui. In particolare, il genitore più a destra elabora due gettoni e li spedisce ai "figli"; mentre il genitore di sinistra non spedisce nulla, poiché la regola vuole che in risposta a un gettone assorbito attraverso un certo lato si elabori un nuovo gettone da spedire sui lati rimanenti;
- (d) un nodo riceve un gettone da un lato, ne elabora uno nuovo e lo spedisce sull'altro lato.

Con quest'ultima fase la rete ha raggiunto un nuovo stato di equilibrio, in cui, a causa delle elaborazioni compiute, i nodi "*scompenso*" e "*nefrosi*" conterranno le probabilità a posteriori delle corrispondenti variabili ospitate. Questo risultato non

ha richiesto la manipolazione dell'intera distribuzione delle variabili del grafo, ma solo tante semplici manipolazioni "locali".

L'esempio appena mostrato descrive un metodo di propagazione "esatto" proposto in (Pearl, 1986), e rielaborato in (Pearl, 1988), ed è valido su reti *singolarmente-connesse* (niente cicli non-orientati). La propagazione "esatta" dell'incertezza su reti di tipo più generale è, purtroppo, un problema NP-completo (Cooper, 1987). L'algoritmo esatto che si è dimostrato il più efficiente su buona parte delle reti non-singolarmente-connesse converte la RCP in una rete di Markov e poi la decompone in "cliques", cioè insiemi massimali di nodi tali per cui tutti i nodi dell'insieme sono adiacenti fra di loro (Lauritzen, 1988). I cliques vengono infine organizzati secondo un albero, che fornisce il supporto al processo di propagazione mediante scambio di messaggi da clique a clique. Ma questo algoritmo non sfugge alla dura legge, e la sua complessità cresce *esponenzialmente* con il grado di interconnessione della rete, rendendole velocemente intrattabile.

È importante, infine, notare che gli algoritmi appena menzionati si applicano a variabili discrete o discretizzate. Di conseguenza, in presenza di variabili continue sarà necessario prevedere le difficoltà e i problemi connessi con la loro discretizzazione, che può avvenire in fase preliminare alla propagazione (discretizzazione fissa) o durante la stessa (discretizzazione dinamica).

## 7. PROPAGAZIONE MEDIANTE METODI APPROSSIMATI

I problemi computazionali connessi con l'uso di tecniche menzionate nel paragrafo precedente hanno deviato le speranze di molti verso *algoritmi approssimati*, cioè algoritmi di propagazione che offrono una complessità *polinomiale* (invece che esponenziale) in cambio di *minore* precisione e accuratezza. Una delle prime proposte viene da Henrion (1988), che ha descritto una tecnica elementare di *simulazione stocastica*, ovvero di *campionamento*, detta *logic sampling*. Pearl (1987), ha successivamente rielaborato l'idea di Henrion, giungendo a proporre uno schema di simulazione alternativo chiamato *Markov sampling*, che risulta imparentato ad uno precedentemente proposto da Geman *et al* (1984).

Le tecniche di campionamento hanno l'obiettivo di estrarre distribuzioni marginali da una distribuzione congiunta specificata attraverso distribuzioni condizionali. Per esempio, con riferimento alla rete in Fig. 7, potremmo voler stimare, ad esempio  $p(P)$ , cioè la distribuzione marginale della variabile  $P$ , sulla base delle distribuzioni condizionali associate ai nodi della rete, cioè  $p(T)$ ,  $p(C)$ ,  $p(P|T)$ ,  $p(A|P,C)$ . Solitamente si dispone di *evidenza*, cioè del valore noto di alcune delle variabili della rete, ad esempio supponiamo di conoscere il valore  $T=t$  e  $A=a$ , quindi



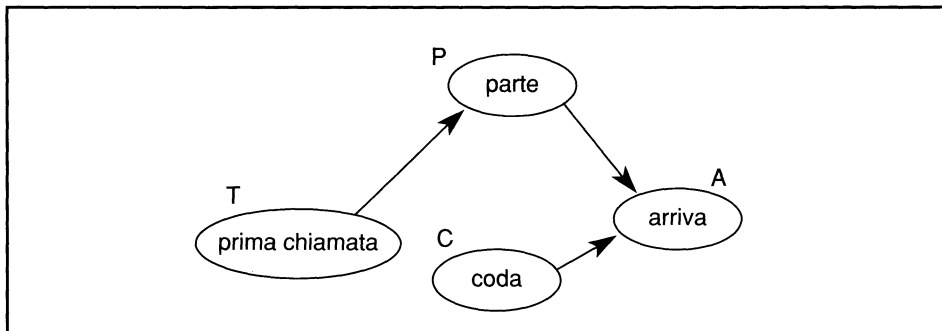


Fig. 7: La rete del nostro “esempio della coda” ridotta.

possiamo scrivere  $EV=\{T=t, A=a\}$ , dove  $EV$  sta per “evidenza”. In questa situazione, il nostro obiettivo potrebbe essere la stima della distribuzione marginale *a posteriori*  $p(P, C | T=t, A=a)$ , ovvero la distribuzione di probabilità congiunta dell’istante della partenza e dell’istante di formazione della coda, dati gli istanti della prima chiamata e dell’arrivo.

L’algoritmo *Markov sampling* “visita” ogni nodo  $X$  su cui non c’è evidenza, e genera mediante campionamento un valore dalla sua distribuzione condizionale rispetto ai valori correnti dei rimanenti nodi (anche quelli con evidenza) della rete. Tale operazione di campionamento, compiuta dal generico nodo  $X$ , si scrive:

$$X \sim p(X|O-X)$$

dove  $O-X$  sta per la totalità dei nodi della rete escluso  $X$ .  $p(X|O-X)$  usualmente ha una forma ridotta  $p(X | N_X)$ , dove  $N_X \subset (O-X)$  denota quell’insieme unico e minimale di nodi *del contorno* di  $X$ , chiamato *frontiera di Markov* di  $X$ , che rende  $X$  indipendente dal resto della rete. La frontiera di Markov di un nodo è formata dai suoi predecessori diretti, dai suoi successori diretti e dai predecessori diretti di tali successori diretti. Per esempio,  $N_p = \{T, C, A\}$  e  $N_c = \{P, A\}$ .

In presenza di variabili continue, si può procedere con un campionamento diretto dalle rispettive distribuzioni continue, oppure risulta talvolta più efficiente eseguire una previa discretizzazione delle variabili continue, per poi procedere campionando da distribuzioni discrete.

Ritorniamo alla rete in figura 7, e supponiamo che si disponga dell’evidenza  $EV=\{T=t, A=a\}$ . Allora, se lo scopo è quello di stimare  $P(P, C | EV)$  mediante Markov sampling genereremo valori  $\pi_i$ , estraendoli dalla distribuzione condizionale  $p(P | T, A, C)$ , e valori  $\chi_i$ , estraendoli dalla distribuzione condizionale  $p(C | P, A)$ . Più precisamente, partendo da un valore iniziale arbitrario ( $\chi_0$ ), l’algoritmo *Markov*

sampling genererebbe una sequenza  $(\pi_i, \chi_i)$ ,  $i = 1, 2, \dots, n$  come segue:

ALGORITMO MARKOV-SAMPLE  $(\chi_0, n)$

For  $i = 1, \dots, n$  do

(i)  $\pi_i \sim p(P=\pi|T=t) p(A=a|P=\pi, C=\chi_{i-1})$

(ii)  $\chi_i \sim p(C=\chi) p(A=a|P=\pi_i, C=\chi)$

End

I teoremi presi in rassegna da Gelfand *et al* (1988) permettono di dimostrare che sotto certe condizioni la sequenza  $(\pi_i, \chi_i)$  è un processo Markoviano ergodico, e converge ad una distribuzione di equilibrio data appunto dalla distribuzione a posteriori cercata, cioè  $p(C, P | T=t, A=a)$ . La convergenza non dipende dall'ordine con cui vengono visitati i nodi, salvo che in ogni nodo deve essere visitato con sufficiente frequenza, ma *può venire distrutta* dalla presenza di stati-pozzo  $(\pi^*, \chi^*)$  dai quali l'algoritmo non può uscire. La presenza di questi stati-pozzo deriva dal fatto che la densità di probabilità sottostante può avere valore nullo in corrispondenza di certi particolari "patterns" di valori su variabili adiacenti.

Si immagini ad esempio che con riferimento alla Figura 5 sia  $P=T+3$ , e lo scopo sia quello di stimare la probabilità marginale a posteriori  $P(P | A=a)$ . In queste condizioni l'algoritmo MARKOV-SAMPLE visita iterativamente i nodi  $T, P$ , e  $C$ . Esso sarà costretto a generare  $\tau_1$  uguale a  $\pi_0-3$ ,  $\pi_1$  uguale a  $\tau_1+3$ ,  $\tau_2$  uguale a  $\pi_1-3$  e così via, di modo che, in pratica, lo stato della rete rimarrà allacciato alla configurazione "pozzo"  $(\tau_0, \pi_0)$ . Il risultato sarà dunque la generazione di una sequenza di campioni identici, tutti uguali allo stato iniziale della rete, che peraltro è arbitrario.

Alla luce di questo problema, si può suggerire un'approccio di simulazione non-iterativo, basato sul principio del *campionamento per importanza* della teoria della simulazione stocastica (vedi, ad esempio Ripley, 1984). Consideriamo ancora, a titolo di esempio, il problema della stima di  $P(P | A=a)$ . Ecco una possibilità: si scelgano distribuzioni di *importanza*  $P^1(T)$  and  $p^1(C)$ , con l'unica condizione che  $P^1(T)$  (risp.  $p^1(C)$ ) abbia supporto positivo dovunque la corrispondente distribuzione  $p(T)$  (risp.  $p(C)$ ) lo ha. Si generino poi terne *i.i.d.*  $(T_i, P_i, C_i)$ ,  $i=1, \dots, N$  estraendo  $T_i$  da  $P^1(T)$ ,  $C_i$  da  $p^1(C)$  e  $P_i$  da  $p(P | T_i, A=a, C_i)$ , cioè da  $p(P | T_i) p(A=a | P, C_i)$ . Si calcolino infine i pesi:

$$r_i = \frac{p(T_i)p(C_i)}{p^1(T_i)p^1(C_i)} \quad i = 1, \dots, N$$

Uno stimatore della distribuzione a posteriori desiderata è il seguente:

$$p^*(P = \pi | A = a) = \frac{\sum_{i=1}^N p(P = \pi | T_i) p(A = a | P = \pi, C_i) r_i}{\sum_{i=1}^N r_i}$$

Questo algoritmo converge lentamente quando nella maggior parte dei campioni generati  $p(A=a|P=\pi, C_i)$  risulta essere vicina a zero. Questo problema può essere superato mediante opportuno “aggiustamento” delle distribuzioni di importanza  $P'(T)$  e  $p'(C)$ , tale da dare a quei campioni che hanno  $p(A=a|P=\pi, C_i)$  alta (i campioni “importanti” appunto) maggiori “chances” di essere estratti. La scelta di  $P'(T)=P(T)$  e  $p'(C)=p(C)$  in genere non è buona. Chiaramente la determinazione di strategie generali per sintetizzare, magari in modo automatico, appropriate distribuzioni di importanza a seconda della particolare rete e dello specifico problema di inferenza richiede ancora un notevole sforzo di ricerca e di sperimentazione numerica.

## 8. CIRCOSCRIZIONE

Nei paragrafi precedenti abbiamo esaminato problemi connessi con l'efficienza degli algoritmi per la propagazione dell'incertezza su reti probabilistiche. Non abbiamo però menzionato una delle soluzioni pratiche più importanti, e cioè la *circoscrizione*. Con questo termine intendiamo qui l'esecuzione di una riduzione della rete preliminare alla propagazione. Più precisamente, data una domanda  $Q$ , prima della propagazione si determina, *senza ricorrere a calcoli numerici*, il sottinsieme dei nodi della rete che può essere scartato senza conseguenze, in quanto cioè non interferisce con  $Q$ . Poi, per generare le probabilità a posteriori che soddisfano la domanda  $Q$  si esegue la propagazione sulla rete formata dai nodi rimanenti.

Per ottenere questo scopo è necessario compiere una trasformazione sulla rete di interesse, e cioè si rappresentano come nodi addizionali le probabilità condizionali associate ai diversi nodi della rete originale. Più precisamente, le probabilità condizionali associate al generico nodo  $X$  della rete originale vengono formalmente rappresentate nella rete trasformata come un ulteriore nodo “genitore” di  $X$ , denotato con il simbolo  $p$ . Ad esempio, la rete in figura 3 si trasforma nella rete in Fig. 8.

Il metodo proposto per la circoscrizione è il seguente: sia  $Y$  l'insieme dei nodi

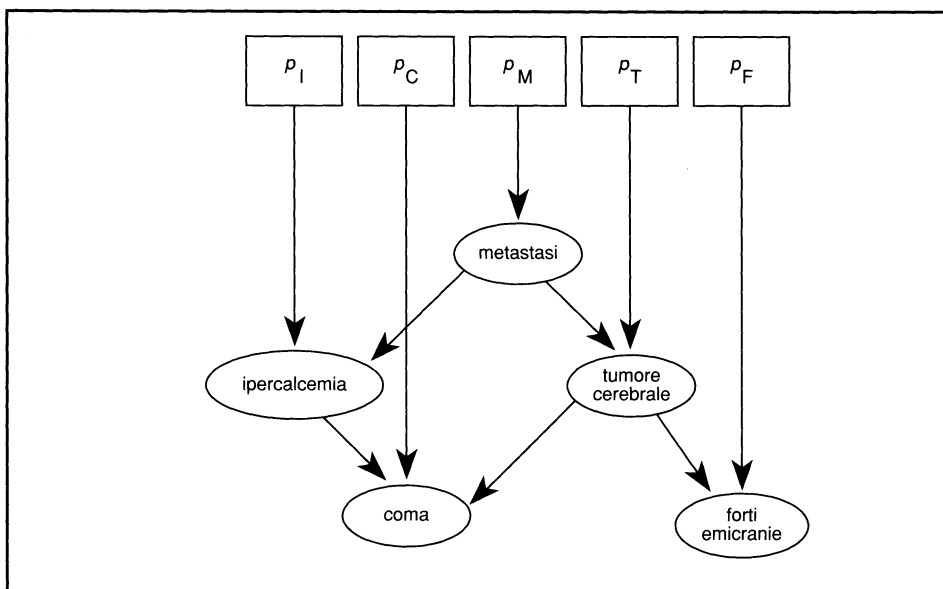


Fig. 8: Le probabilità condizionali associate ai nodi sono formalmente rappresentate come nodi della rete.

che nella rete originale rappresentano ipotesi cui si riferisce la domanda  $Q$ , ed  $E$  l'insieme dei nodi osservati nella rete originale; un nodo *osservato*  $X$  può essere rimosso dalla rete originale se nella rete trasformata valgono contemporaneamente  $i(X, E-X, Y)_D$  e  $i(p_X, E, Y)_D$ , mentre un nodo non osservato  $X$  può essere rimosso dalla rete se vale la  $i(p_X, E, Y)_D$ .

Vediamo un esempio applicativo: con riferimento alla rete in Fig. 8, supponiamo che si voglia rispondere alla domanda “in un paziente con precedente storia di *metastasi* e sofferente di *ipercalcemia*, qual è la probabilità che in futuro si sviluppi un *tumore cerebrale*?”. La forma con cui è posta la domanda implica la non disponibilità di informazione su *coma* e *forti-emicranie*, poiché queste variabili si riferiscono a un'epoca temporale successiva a *tumore-cerebrale*. In tal caso, l'applicazione del criterio sopra riportato porta a ridurre la rete originale (quella in Fig. 3) alla sola coppia di nodi *metastasi* e *tumore-cerebrale* (e il lato che li congiunge), scartando anche l'informazione circa l'*ipercalcemia* come irrilevante.

Supponiamo ora di cambiare la domanda in “dato un paziente con precedente storia di *metastasi*, sofferente di *ipercalcemia* e attualmente in *coma*, qual è la probabilità che soffra di un *tumore-cerebrale*?”. L'informazione riguardante “coma” attiva un cammino fra “ipercalcemia” e “tumore cerebrale”, per cui il nodo *ipercalcemia* non può essere eliminato. L'applicazione del criterio porta a concludere che solo

il nodo *forti-emicranie* (se tale informazione non è disponibile) può essere eliminato. L'inclusione del nodo *ipercalcemia* è intuitiva, poiché tenendo conto che il paziente soffre di *ipercalcemia* possiamo sperare che il coma sia un prodotto di tale disturbo e, conseguentemente, ridurre il nostro timore a *tumore cerebrale*.

### RIFERIMENTI BIBLIOGRAFICI

- Beeri, C., Fagin R., Maier, D., e Yannakakis, M., 1983, On the desirability of acyclic data-base schemes, *JACM*, 30, (no. 3), 479–513.
- Berzini C., 1989, A probabilistic framework for temporal reasoning, accettato per pubblicazione sulla rivista *Artificial Intelligence*.
- Blalock, H.M., 1971, *Causal Models in the Social Sciences*, MacMillan, London.
- Clayton, D.G., 1988, The analysis of event history data: a review of progress and outstanding problems, *Statistics in Medicine*, 7, 819–841.
- Cox, R., 1946, Probability frequency and reasonable expectation. *American Journal of Physics*, 14 (n° 1), 1–13.
- de Finetti, B., 1937, Foresight: its logical laws, its subjective sources, *Ann. Inst. H. Poincaré*, 7 1–68. Tradotto e ristampato in (KYBURG 80).
- Duda, R.O., Hart, P.E., e Nilsson, N.J., 1976, Subjective Bayesian methods for rule-based inference systems, *Proceedings Natl. Comp. Conf. (AFIPS)*, 45, 1075–82.
- Duncan, O.D., 1975, *Introduction to Structural Equation Models*, Academic Press, New York.
- Gelfand, A.E., & Smith, A.F.M., 1988 Sampling based approaches to calculating marginal densities, *Technical Report*, Departement of Mathematics, University of Nottingham, U.K.
- Geman, S & Geman, D., 1984, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721.
- Goodman, L.A., 1970, The multivariate analysis of quantitative data: Interaction among multiple classifications, *J., Amer. Statist. Assoc.*, 65, 226–56.
- Haberman, S.J., 1974, The genral log-linear model, Ph.D. diss., Department of Statistics, University of Chicago.
- Henrion, M., 1988, Propagating Uncertainty in Bayesian Networks by Probabilistic Logic Sampling, in: *Uncertainty in Artificial Intelligence 2*, eds, L.N. Kanal and J.F. Lemmer (North Holland, 1988) p.149.
- Howard, R.A., and Matheson, J.E., 1981, Influence diagrams, in: *Principles and applications of decision analysis*, vol.2 (1984). Menlo Park, Calif.: Strategic Decision Group.
- Kyburg, H.E., Smokler, H.E. (eds.), 1980, *Studies in subjective probability*, Krieger.
- Lauritzen, S.L., 1982, *Lectures on Contingency Tables*, University of Aalborg Press.
- Lauritzen, S.L., Spiegelhalter, D.J., 1988, Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems, *J.R. Statist. Soc B*, 50, No. 2, 157.
- Pearl, J., 1986, Fusion, propagation, and structuring of belief networks. *Artificial Intelligence*, 29, 241.
- Pearl, J., 1987, Evidential reasoning using stochastic simulation of casual models, *Artificial Intelligence*, 32, no. 2, 245.
- Pearl, J., 1988, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo.
- Ramsey, F.P., 1931, Truth and Probability, in: *The foundations of mathematics and other logical essays* (R.B. Braithwaite ed.). New York: Humanities Press. Ristampato in (KYBURG 80).

- Ripley, B.D., 1987 *Stochastic Simulation*, Wiley & Sons.
- Shacter, R.D., 1986, Evaluating Influence Diagrams, *Operations Research*, 34 (no. 6) 871–82.
- Smith, J.Q., 1987, Influence diagrams for statistical modeling, Technical Report 117, Department of Statistics, University of Warwick, U.K.
- Wold, H, 1964, *Econometric model building*, North-Holland, Amsterdam.
- Wright, S., 1934, The method of path coefficients, *Ann. Math. Statistics*, 5, 161–215.

### SUMMARY

*Computational models of human reasoning based on the axioms of probability theory are discussed, with applications to "expert systems". A central concept is the possibility of representing probabilistic information via graphical models, or networks. The two aspects of a probabilistic model, namely the structure of dependencies among variables, and the probability specifications can be separately represented. The former aspect, intrinsically qualitative, is expressed by the connectivity structure of a directed acyclic graph, while the latter is expressed through conditional probability tables stored in the nodes of the network. Such an organization allows for parallel and distributed computation, and allows reasoning activity concerning the represented variables to be vividly displayed by tracing pathways along the network. This is an overview of recent research result concerning (i) criteria for identifying independencies implied by the topology of a probabilistic network, (ii) the limits of the expressiveness of a probabilistic network, and (iii) algorithms for propagating probabilities over a graphical model.*

*Parole chiave: sistemi esperti probabilistici, reti casuali probabilistiche, ragionamento temporale, processi stocastici, modelli bayesiani.*