Incertezza

Capitolo 13

Outline

◊ Incertezza

◊ Probabilità

◊ Sintassi e semantica

◊ Inferenza

◊ Indipendenza e Regola di Bayes

Incertezza

Azione A_t = partire per l'aeroporto t minuti prima del volo
L'azione A_t mi permetterà di arrivare in tempo?

Problemi:
1) osservabilità parziale (stato della strada, piano di altri veicoli, etc.)
2) sensori rumorosi (rapporti sul traffico di Isoradio ...)
3) incertezza nell'esito delle azioni (pneumatico forato, etc.)
4) immensa complessità nel modellare e nel predire il traffico

Quindi un approccio puramente logico o
1) rischia di dire il falso: "A_{25} mi fa arrivare in tempo"
o 2) conduce a conclusioni che sono troppo deboli per prendere decisioni:
"A_{25} mi fa arrivare in tempo se non c''e un incidente sul ponte
e non piove e non foro i pneumatici, etc etc."

(A_{1440} si può ragionevolmente dire che mi fa arrivare in tempo
ma devo passare la notte all'aeroporto ...)

Metodi per trattare l'incertezza

Default o nonmonotonic logic:
Assumiamo che l'auto non abbia forato
Assumiamo che A_{25} vada bene a meno che l'evidenza non lo contraddica
Problemi: quali assunzioni sono ragionevoli? Trattamento contraddizioni?

Regole con fattori di "inganno" (fudge factors):
$A_{25} \mapsto_{0.3}$ mi fa arrivare in tempo
$irrigatore \mapsto_{0.99} erba-bagnata$
$erba-bagnata \mapsto_{0.7} pioggia$
Problemi: come trattare la combinazione ? p.e., $irrigatore$ causa $pioggia$??

Probabilità
Data l'evidenza disponibile,
A_{25} mi fa arrivare in tempo con probabilità 0.04
Mahaviracarya (9th C.), Cardamo (1565) teoria del gioco

(Logica Fuzzy manipola *gradi di verità* NON incertezza p.e.,
$erba-bagnata$ è vera con grado 0.2)

Probabilità

Asserzioni Probabilistiche *riassumono* gli effetti di
- pigrizia: fallimento nell'enumerare le eccezioni, qualifica, etc.
- ignoranza: mancanza di fatti rilevanti, condizioni iniziali, etc.

Probabilità Soggettiva o Bayesiana:
Le probabilità legano le proposizioni al proprio stato di conoscenza
p.e., $P(A_{25}|\text{nessun incidente riportato}) = 0.06$

Non indica alcuna tendenza probabilistica nella situazione corrente

Le probabilità delle proposizioni cambiano con l'arrivo di nuova evidenza:
p.e., $P(A_{25}|\text{nessun incidente riportato, 5 a.m.}) = 0.15$

Decidere nell'incertezza

Supponiamo di credere che:

$$\begin{aligned}
P(A_{25}\text{ mi fa arrivare in tempo}|\ldots) &= 0.04\\
P(A_{90}\text{ mi fa arrivare in tempo}|\ldots) &= 0.70\\
P(A_{120}\text{ mi fa arrivare in tempo}|\ldots) &= 0.95\\
P(A_{1440}\text{ mi fa arrivare in tempo}|\ldots) &= 0.9999
\end{aligned}$$

Quale azione scegliere ?

Dipende dalle mie preferenze: perdere l'aereo vs. la cucina dell'areoporto, etc.

Teoria dell'utilità è usata per rappresentare e inferire preferenze

Teoria delle decisioni = teoria dell'utilità + teoria delle probabilità

Elementi di Probabilità

Insieme Ω— *spazio degli eventi*
- p.e., 6 possibili lanci di dado.
- $\omega \in \Omega$ è un campione (possibile evento atomico)

Uno *spazio di probabilità* o *modello di probabilità* è uno spazio degli eventi con un assegnamento $P(\omega)$ per ogni $\omega \in \Omega$ tale che
- $0 \leq P(\omega) \leq 1$
- $\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

Un *evento* A è un qualunque sottoinsieme di Ω

$$P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$$

P.e., $P(\text{lancio dado} < 4) = 1/6 + 1/6 + 1/6 = 1/2$

Variabili aleatorie

Una *variabile aleatoria* è una funzione da eventi a qualche intervallo, p.e., i reali o i Booleani
p.e., $Dispari(1) = vero$.

P induce una *distribuzione di probabilità* per ogni variabile aleatoria X:

$$P(X = x_i) = \Sigma_{\{\omega : X(\omega) = x_i\}} P(\omega)$$

p.e., $P(Dispari = vero) = 1/6 + 1/6 + 1/6 = 1/2$

Proposizioni

Pensare ad una proposizione come un evento (insieme di eventi) dove la proposizione è vera

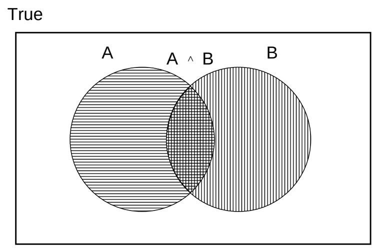
Date variabili Booleane aleatorie A e B:
  evento a = insieme di eventi dove $A(\omega) = vero$
  evento $\neg a$ = insieme di eventi dove $A(\omega) = falso$
  evento $a \wedge b$ = eventi dove $A(\omega) = vero$ e $B(\omega) = vero$

Variabili Booleane, evento = modello di logica proposizionale
    p.e., $A = vero$, $B = falso$, o $a \wedge \neg b$.
Proposizione = disgiunzione di eventi atomici in cui essa è vera
  p.e., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
  $\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Perché usare le probabilità ?

Le definizioni implicano che alcuni eventi logicamente correlati devono avere probabilità correlate

P.e., $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



de Finetti (1931): un agente che scommette secondo probabilità che violano questi assiomi possono essere forzati a perdere soldi indipendentemente dall'esito degli eventi

Sintassi per le proposizioni

Variabili aleatorie proposizionali o Booleane
  p.e., *Cavità* (ho una cavità ?)

Variabili aleatorie discrete (*finite* o *infinite*)
  p.e., $Tempo$ può assumere i valori $\langle sole, pioggia, nuvole, neve\rangle$
  $Tempo = pioggia$ è una proposizione
  I valori devono essere esaustivi e mutuamente escusivi

Variabili aleatorie continue
  p.e., $Temperatura = 21.6$; ammette anche, p.e., $Temperatura < 22.0$.

Combinazioni Booleane arbitrarie di proposizioni base

Probabilità a priori

Prior o probabilità incondizionate di proposizioni
  p.e., $P(\textit{Cavità} = vero) = 0.1$ e $P(Tempo = sole) = 0.72$
corrispondono a gradi di credenza sull'arrivo di (nuova) evidenza

Distribuzione di probabilità fornisce valori per tutti i possibili assegnamenti:
  $\mathbf{P}(Tempo) = \langle 0.72, 0.1, 0.08, 0.1\rangle$ (*normalizzate*, cioè, sommano ad 1)

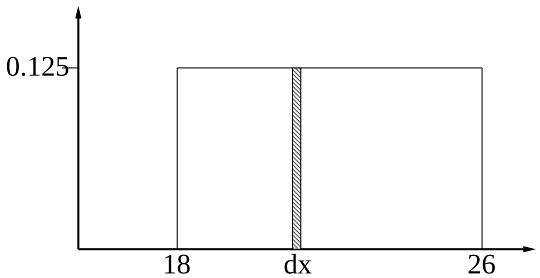
Distribuzione di probabilità congiunta per un insieme di v.a. fornisce la probabilità per ogni evento atomico su tali variabili
  $\mathbf{P}(Tempo, \textit{Cavità})$ = una matrice 4×2 di valori:

$Tempo =$	$sole$	$pioggia$	$nuvole$	$neve$
$\textit{Cavità} = vero$	0.144	0.02	0.016	0.02
$\textit{Cavità} = falso$	0.576	0.08	0.064	0.08

Ogni domanda che concerne un dominio trova risposta nella distribuzione congiunta perché ogni evento è la somma dei possibili eventi

Probabilità per variabili continue

Esprimere la distribuzione come una funzione parametrizzata di valori:
$P(X = x) = U[18, 26](x) =$ densità uniforme tra 18 e 26

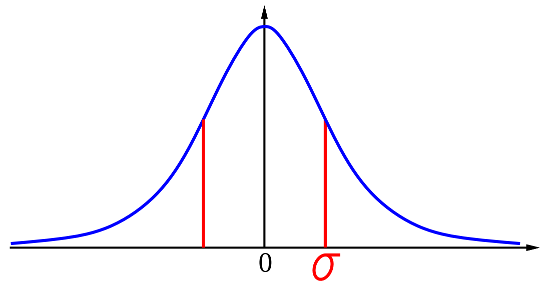


P è una *densità*; integra ad 1.
$P(X = 20.5) = 0.125$ in realtà significa

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$$

Densità Gaussiana

$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$



Probabilità condizionale

Probabilità a posteriori o condizionale
p.e., $P(\textit{Cavità}|Mal\_di\_denti) = 0.8$
cioè, dato che Mal_di_denti è tutto quello che so

Se so di più, p.e., *Cavità* è anche data, allora abbiamo
$P(\textit{Cavità}|Mal\_di\_denti, \textit{Cavità}) = 1$
Nota: la credenza meno specifica *rimane valida* dopo che nuova evidenza arriva, ma non necessariamente rimane *utile*

Nuova evidenza può essere irrilevante, permettendo semplificazioni, p.e.,
$P(\textit{Cavità}|Mal\_di\_denti, Vince\_Inter) = P(\textit{Cavità}|Mal\_di\_denti) = 0.8$
Questo tipo di inferenza, dovuta alla conoscenza del dominio, è cruciale

Probabilità condizionale

Definizione di probabilità condizionale:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

La regola del prodotto fornisce una definizione alternativa:
$P(a \wedge b) = P(a|b)P(b) = p(b|a)P(a)$

Una versione generale vale sulle distribuzioni, p.e.,
$\mathbf{P}(Tempo, \textit{Cavità}) = \mathbf{P}(Tempo|\textit{Cavità})\mathbf{P}(\textit{Cavità})$
(Visto come un insieme 4×2 di equazioni, *no* moltiplicazione di matrici)

Chain rule è derivata dalla applicazione ripetuta della regola del prodotto:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n_1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \Pi_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})\end{aligned}$$

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Per ogni proposizione ϕ, si sommano gli eventi atomici dove essa è vera:
$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Per ogni proposizione ϕ, si sommano gli eventi atomici dove essa è vera:
$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Per ogni proposizione ϕ, si sommano gli eventi atomici dove essa è vera:
$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

$P(cavity \vee toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

Inferenza tramite enumerazione

Si inizia con la distribuzione congiunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Si possono calcolare anche le probabilità condizionali:

$$\begin{aligned} P(\neg cavity | toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 \end{aligned}$$

Normalizzazione

	toothache		¬ toothache	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

Il denominatore può essere visto come una *costante di normalizzazione* α

$$\begin{aligned}
&\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)\\
&\quad = \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]\\
&\quad = \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]\\
&\quad = \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle
\end{aligned}$$

Idea generale: calcolare la distribuzione sulla variabile della query fissando le variabili di evidenza e sommando sulle variabili nascoste

Inferenza tramite enumerazione

Tipicamente siamo interessati a
la distribuzione congiunta a posteriori delle variabili di query $\mathbf{Y}$
dati specifici valori e per le variabili di evidenza $\mathbf{E}$

Poniamo le variabili nascoste essere $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Allora la somma desiderata di entrate congiunte è ottenuta sommando sulle variabili nascoste:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha\Sigma_{\mathbf{h}}\mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$$

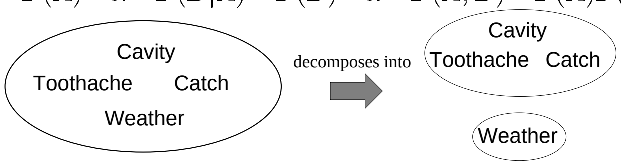
I termini nella sommatoria sono entrate congiunte perché $\mathbf{Y}$, $\mathbf{E}$, e $\mathbf{H}$ insieme esauriscono l'insieme delle variabili aleatorie

Problemi ovvi:

1) Complessità caso pessimo in tempo $O(d^n)$ dove d è l'arietà più grande
2) Complessità in spazio $O(d^n)$ per memorizzare la distribuzione congiunta
3) Come stabilire i valori per $O(d^n)$ entrate???

Indipendenza

A e B sono indipendenti sse
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



$\mathbf{P}(Toothache, Catch, Cavity, Weather)$
$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$

32 entrate ridotte a 12; per n monete "truccate" indipendenti, $2^n \to n$

Indipendenza assoluta potente ma rara

Nei problemi reali sono coinvolte centinaia di variabili, nessuna delle quali è indipendente. Che fare?

Indipendenza condizionale

$\mathbf{P}(Toothache, Cavity, Catch)$ ha $2^3 - 1 = 7$ entrate indipendenti

Se si ha una cavità, la probabilità che la sonda si fermi su in essa non dipende dal fatto di avere il mal di denti:
(1) $P(catch|toothache, cavity) = P(catch|cavity)$

La stessa indipendenza vale se non c'è la cavità:
(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

Catch è *condizionalmente indipendente* da *Toothache* dato *Cavity*:
$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$

Affermazione equivalente:
$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$
$\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$

Indipendenza condizionale

Scrivere la distribuzione congiunta completa usando la chain rule:

$$\begin{aligned}
&\mathbf{P}(Toothache, Catch, Cavity)\\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity)\\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)\\
&= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

Cioè, $2 + 2 + 1 = 5$ numeri indipendenti

In molti casi, l'uso di indipendenza condizionale riduce la dimensione della rappresentazione della probabilità congiunta da essere esponenziale in n a lineare n.

L'indipendenza condizionale è la forma più basilare e robusta di conoscenza sugli ambienti incerti.

Regola di Bayes

Regola del prodotto $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$$\Rightarrow \text{Bayes' rule } P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

o in forma di distribuzione

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha\mathbf{P}(X|Y)\mathbf{P}(Y)$$

Utile per ottenere probabilità diagnostica a partire da probabilità causale:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

P.e., sia M la rappresentazione di Meningite, e S di collo rigido:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

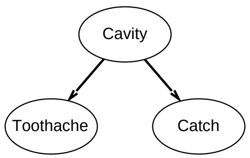
Nota: la probabilità a posteriori della Meningite ancora piccola!

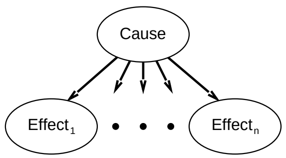
Regola di Bayes e indipendenza condizionale

$$\begin{aligned}
&\mathbf{P}(Cavity|toothache \wedge catch)\\
&= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)\\
&= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

Questo è un esempio di modello *naive Bayes*:

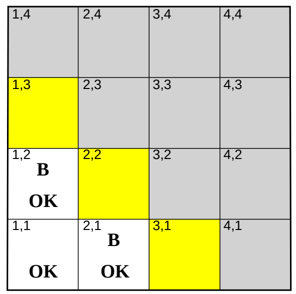
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i\mathbf{P}(Effect_i|Cause)$$





Il numero totale di parametri è *lineare* in n

Mondo dei Wumpus



$P_{ij} = vero$ sse $[i, j]$ contiene una trappola

$B_{ij} = vero$ sse $[i, j]$ è ventilato
Includiamo solo $B_{1,1}, B_{1,2}, B_{2,1}$ nel modello probabilistico

Specifica del modello probabilistico

La distribuzione congiunta completa è $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Applicare la regola del prodotto: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$

(facciamo così per ottenere $P(Effect|Cause)$.)

Primo termine: 1 se le trappole sono adiacenti a "brezze", 0 altrimenti

Secondo termine: le trappole sono posizionate a caso, con probabilità 0.2 per quadrato:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{i,j=1,1}^{4,4}\mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

per n trappole.

Osservazioni e query

Noi conosciamo i seguenti fatti:

$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$
$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$

La query è $\mathbf{P}(P_{1,3}|known, b)$

Definiamo $unkown =$ i P_{ij} diversi da $P_{1,3}$ e $known$

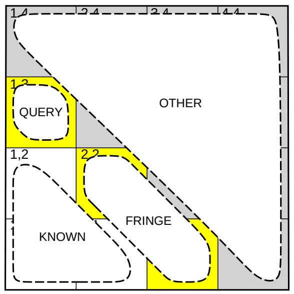
Per effettuare inferenza per enumerazione, abbiamo

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \Sigma_{unkown} \mathbf{P}(P_{1,3}, unkown, known, b)$$

Cresce esponenzialmente con il numero di quadrati!

Usando l'indipendenza condizionale

Idea base: le osservazioni sono condizionalmente indipendenti da altri quadrati nascosti dati i quadrati nascosti adiacenti



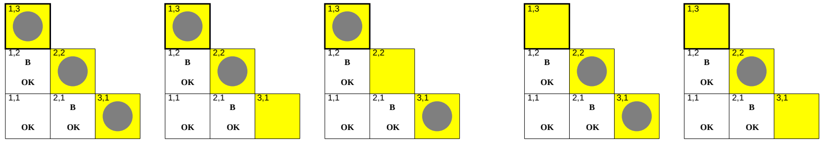
Definiamo $Unkown = Fringe \cup Other$
$\mathbf{P}(b|P_{1,3}, Known, Unkown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$

Poniamo la query in una forma dove si possa usare quanto sopra!

Usando l'indipendenza condizionale

$$\begin{aligned}
\mathbf{P}(P_{1,3}|known, b) &= \alpha \sum_{unkown} \mathbf{P}(P_{1,3}, unkown, known, b) \\
&= \alpha \sum_{unkown} \mathbf{P}(b|P_{1,3}, known, unkown)\mathbf{P}(P_{1,3}, known, unkown) \\
&= \alpha \sum_{fringe}\sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe}\sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other) \\
&= \alpha\, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other) \\
&= \alpha'\, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)
\end{aligned}$$

Usando l'indipendenza condizionale



0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16 | 0.8 x 0.2 = 0.16 | 0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16

$$\begin{aligned}\mathbf{P}(P_{1,3}|known, b) &= \alpha' \langle 0.2(0.04 + 0.16 + 0.16),\ 0.8(0.04 + 0.16) \rangle \\ &\approx \langle 0.31, 0.69 \rangle\end{aligned}$$

$$\mathbf{P}(P_{2,2}|known, b) \approx \langle 0.86, 0.14 \rangle$$

Riassunto

Il calcolo delle probabilità costituisce un formalismo rigoroso per la conoscenza incerta

La distribuzione congiunta di probabilità specifica la probabilità di ogni evento atomico

Si può rispondere alle query sommando sugli eventi atomici

Per domini non banali, bisogna trovare un modo per ridurre la dimensione della rappresentazione della probabilità congiunta

Indepenclenza ed indipendenza condizionale forniscono tale modo