

Sommario

Esercitazione Richiami di probabilità.....	1
Reti Bayesiane e D-Separation - ASSIGNMENT.....	7
Esercitazione Reti Bayesiane, distribuzione di probabilità congiunta e indipendenza.....	10
Inferenza Esatta e Approssimata - ASSIGNMENT	13
Esercitazione Inferenza Esatta e Approssimata	22
Catene di Markov - ASSIGNMENT.....	26
Esercitazione Catene di Markov	30
Hidden Markov Model - ASSIGNMENT.....	34
Esercitazioni Hidden Makrov Model.....	37
Esercizi preparazione Esame	Errore. Il segnalibro non è definito.
TEORIA	46
Rete Bayesiana e D-Separazione	46
Inferenza esatta.....	47
Generazione numeri pseudo-casuali.....	48
Inferenza approssimata.....	49
Hidden Markov Models – Catene di Markov.....	50
Hidden Markov Models – Inferenza	52
Filtri di Kalman	56

Esercitazione Richiami di probabilità

Esercizio 1. Si supponga che il 5% degli uomini e lo 0.25% delle donne siano daltonici.
Supponiamo all'inizio che vi sia lo stesso numero di uomini e di donne.

1. Qual è la percentuale di daltonici (uomini e donne) nella popolazione totale?
2. Se si sceglie a caso un daltonico, qual è la probabilità che questo sia un uomo?

Supponiamo ora che gli uomini siano il doppio delle donne.

3. Qual è la percentuale di daltonici nella popolazione totale?
4. Se si sceglie a caso un daltonico, qual è la probabilità che stavolta sia una donna?

Esercizio 1. Definiamo gli eventi

$$D := \{\text{daltonico}\}, \quad M := \{\text{uomo}\}, \quad F := \{\text{donna}\},$$

dove ovviamente abbiamo $M \cup F = \Omega$, $M \cap F = \emptyset$. Allora i dati si possono riscrivere così:

$$\mathbb{P}(D|M) = 0.05, \quad \mathbb{P}(D|F) = 0.0025.$$

1. Abbiamo come ipotesi che $\mathbb{P}(M) = \mathbb{P}(F) = 0.5$. Allora usando la formula della probabilità totale calcoliamo

$$\mathbb{P}(D) = \mathbb{P}(D|M)\mathbb{P}(M) + \mathbb{P}(D|F)\mathbb{P}(F) = 0.05 \cdot 0.5 + 0.0025 \cdot 0.5 = 0.02625 = \frac{21}{800}$$

2. Usando la formula di Bayes, la probabilità richiesta è:

$$\mathbb{P}(M|D) = \frac{\mathbb{P}(D|M)\mathbb{P}(M)}{\mathbb{P}(D)} = \frac{0.05 \cdot 0.5}{0.02625} = 0.9523 = \frac{20}{21}$$

3. Stavolta abbiamo come ipotesi che $\mathbb{P}(M) = \frac{2}{3}$, $\mathbb{P}(F) = \frac{1}{3}$. Allora usando la formula della probabilità totale calcoliamo

$$\mathbb{P}(D) = \mathbb{P}(D|M)\mathbb{P}(M) + \mathbb{P}(D|F)\mathbb{P}(F) = 0.05 \cdot 0.666 + 0.0025 \cdot 0.333 = 0.03416 = \frac{41}{1200}$$

4. Usando la formula di Bayes, la probabilità richiesta stavolta è:

$$\mathbb{P}(F|D) = \frac{\mathbb{P}(D|F)\mathbb{P}(F)}{\mathbb{P}(D)} = \frac{0.0025 \cdot 0.333}{0.03416} = 0.02439 = \frac{1}{41}$$

Una compagnia di assicurazioni di automobili divide i guidatori in 3 categorie, divise per età: gruppo A (sotto i 25 anni, 17% dei suoi assicurati), gruppo B (25-39 anni, 45% dei suoi assicurati), gruppo C (dai 40 anni in su). I dati mostrano che ogni anno in media le percentuali di assicurati che hanno un incidente sono 13% per il gruppo A, 4% per il gruppo B e 3% per il gruppo C.

Si calcoli la probabilità che un assicurato che ha appena avuto un incidente ha meno di 25 anni.

Si riporti lo svolgimento sul foglio protocollo e si riporti solamente la risposta all'esercizio nel form sottostante.

utilizzeremo il teorema di Bayes. Il teorema di Bayes si basa sulla formula:

$$\mathbb{P}(A|B) = \mathbb{P}(B|A) * \mathbb{P}(A) / \mathbb{P}(B)$$

Dove:

- $\mathbb{P}(A|B)$ è la probabilità che A sia vero dato che B è vero.
- $\mathbb{P}(B|A)$ è la probabilità che B sia vero dato che A è vero.
- $\mathbb{P}(A)$ è la probabilità che A sia vero.
- $\mathbb{P}(B)$ è la probabilità che B sia vero.

In questo esercizio, vogliamo calcolare la probabilità che un assicurato che ha appena avuto un incidente abbia meno di 25 anni, quindi:

- A: un assicurato ha meno di 25 anni (gruppo A)
- B: un assicurato ha avuto un incidente

Per prima cosa, calcoliamo la probabilità di avere un incidente per ciascun gruppo età:

1. $P(B|A) = 13\%$ (percentuale di incidenti per il gruppo A)
2. $P(B|B) = 4\%$ (percentuale di incidenti per il gruppo B)
3. $P(B|C) = 3\%$ (percentuale di incidenti per il gruppo C)

Ora calcoliamo la probabilità per ciascun gruppo età:

1. $P(A) = 17\%$ (percentuale di assicurati nel gruppo A)
2. $P(B) = 45\%$ (percentuale di assicurati nel gruppo B)
3. $P(C) = 100\% - (17\% + 45\%) = 38\%$ (percentuale di assicurati nel gruppo C)

Per calcolare la probabilità che un assicurato che ha appena avuto un incidente sia nel gruppo A, dobbiamo prima calcolare la **probabilità complessiva** di avere un incidente ($P(B)$):

$$P(I) = P(I|A) * P(A) + P(I|B) * P(B) + P(I|C) * P(C)$$

$$P(B) = 0.13 * 0.17 + 0.04 * 0.45 + 0.03 * 0.38$$

$$P(B) \approx 0.0608$$

Ora possiamo applicare il teorema di **Bayes** per calcolare $P(A|B)$:

$$P(A|B) = P(B|A) * P(A) / P(B)$$

$$P(A|B) = (0.13 * 0.17) / 0.0608$$

$$P(A|B) \approx 0.3649$$

La probabilità che un assicurato che ha appena avuto un incidente abbia meno di 25 anni è circa 36.49%.

Esercizio 1

- Dimostrare che $P(a|b \wedge a) = 1$.

$$P(a|b \wedge a) = \frac{P(a \wedge b \wedge a)}{P(b \wedge a)} = \frac{P(a \wedge b)}{P(b \wedge a)} = 1$$

NON SERVE

Esercizio 2

- Date le seguenti belief di un agente razionale

$$P(A) = 0.4 \quad P(B) = 0.3 \quad P(A \vee B) = 0.5$$

quale range di probabilità è ragionevole per $A \cap B$?

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B) =$$

$$= 0.4 + 0.3 - P(A \wedge B) = 0.5$$

$$P(A \wedge B) = 0.2$$

NON SERVE

Esercizio 3

- Consideriamo l'insieme di tutte le possibili mani di una partita a poker con 5 carte, utilizzando un mazzo formato da 52 carte.

1. Quanti eventi atomici costituiscono la distribuzione di probabilità congiunta?
2. Qual è la probabilità di ciascun evento atomico?
3. Qual è la probabilità di avere una scala reale?
4. E di un poker (4 carte dello stesso tipo)?

NON HO CAPITO E NON SERVE

Esercizio 3

3. Qual è la probabilità di avere una scala reale?

1. Quanti eventi atomici costituiscono la distribuzione di probabilità congiunta?

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 2,598,960$$

$$\frac{4}{2,598,960} = \frac{1}{649,740}$$

4. E di un poker (4 carte dello stesso tipo)?

2. Qual è la probabilità di ciascun evento atomico?

$$\frac{1}{2,598,960}$$

$$\frac{13 \cdot 48}{2,598,960} = \frac{1}{4,165}$$

Esercizio 4

- Una società di consulenza ha creato un modello per prevedere le recessioni. Il modello prevede una recessione con l'80% di probabilità quando la recessione avviene realmente e con il 10% di probabilità quando non avviene. La probabilità incondizionata che si entri in una fase di recessione è del 20%. Se il modello prevede la recessione, qual è la probabilità che la recessione avvenga?

BAYES

Il modello prevede una recessione con l'80% di probabilità quando la recessione avviene realmente

$$P(\text{rec. pred.} | \text{rec. coming}) = \frac{8}{10}$$

Il modello prevede una recessione il 10% di probabilità quando non avviene

$$P(\text{rec. pred.} | \text{rec. not coming}) = \frac{1}{10}$$

La probabilità incondizionata che si entri in una fase di recessione è del 20%.

$$P(\text{rec. coming}) = \frac{2}{10}$$

$$P(\text{rec. not coming}) = 1 - P(\text{rec. coming}) = 1 - \frac{2}{10} = \frac{8}{10}$$

- Se il modello prevede una recessione, qual è la probabilità che la recessione effettivamente avvenga?

$$\begin{aligned} & P(\text{rec. coming} | \text{rec. pred.}) = \frac{8}{10} \\ & P(\text{rec. pred.} | \text{rec. coming}) = \frac{2}{10} \\ & P(\text{rec. pred.}) = ? \\ & P(\text{rec. pred.}) = \frac{8}{10} \cdot \frac{2}{10} + \frac{1}{10} \cdot \frac{8}{10} = \frac{24}{100} = \frac{24}{100} = \frac{2}{3} \end{aligned}$$

$$\begin{aligned} P(\text{rec. pred.}) &= P(\text{rec. pred.} | \text{rec. coming})P(\text{rec. coming}) \\ &\quad + P(\text{rec. pred.} | \text{rec. not coming})P(\text{rec. not coming}) \\ &= \frac{8}{10} \cdot \frac{2}{10} + \frac{1}{10} \cdot \frac{8}{10} \\ &= \frac{24}{100} \end{aligned}$$

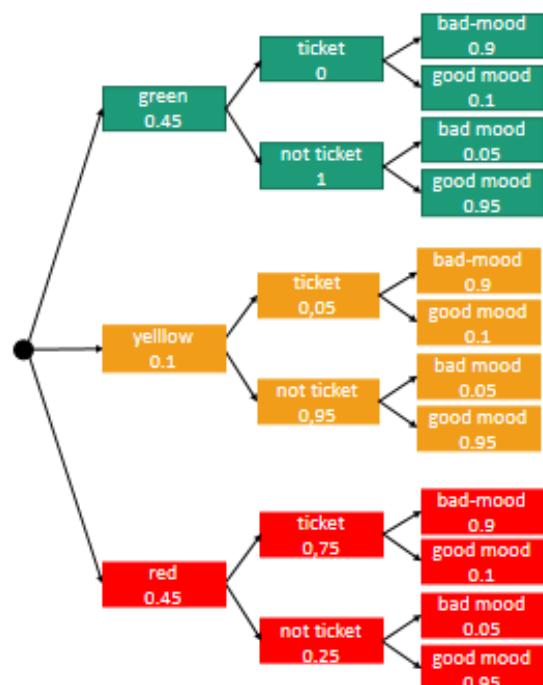
BAYES

Esercizio 5

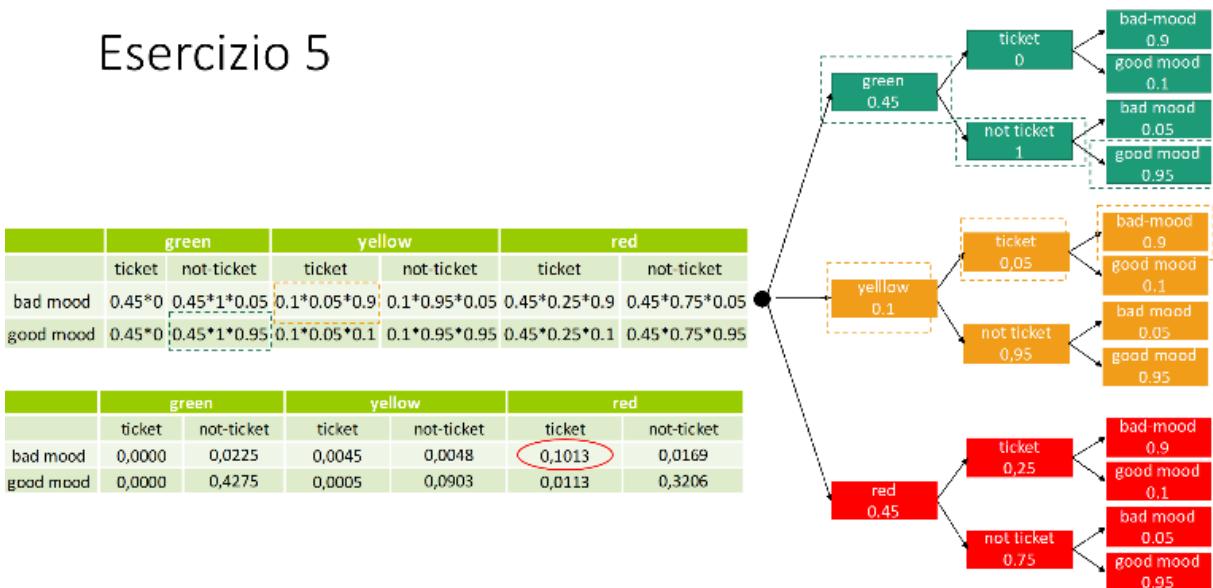
- Supponiamo di conoscere che la probabilità con cui un semaforo diventi verde sia pari a 0.45, arancione pari a 0.1 e che diventi rosso sia pari a 0.45.
- Inoltre, supponiamo di avere la probabilità del 25% di passare con il semaforo rosso senza prendere una multa , e il 5% di probabilità di prendere una multa passando con il semaforo arancione.
- In aggiunta, supponiamo che prendendo una multa, c'è il 90% di probabilità che si sia successivamente di cattivo umore; senza multa la probabilità è del 5%.
- Qual è la probabilità totale di essere di cattivo umore, prendendo una multa, passando con il semaforo rosto?

Esercizio 5

- Supponiamo di conoscere che la probabilità con cui un semaforo diventi verde sia pari a 0.45, arancione pari a 0.1 e che diventi rosso sia pari a 0.45.
- Inoltre, supponiamo di avere la probabilità del 25% di passare con il semaforo rosso senza prendere una multa , e il 5% di probabilità di prendere una multa passando con il semaforo arancione.
- In aggiunta, supponiamo che prendendo una multa, c'è il 90% di probabilità che si sia successivamente di cattivo umore; senza multa la probabilità è del 5%.



Esercizio 5



Formula probabilità congiunta $P(X,Y) = P(X|Y)*P(Y)$.

$P(\text{bad mood, ticket, red}) =$

$= P(\text{bad mood} | \text{ticket, red}) P(\text{ticket, red}) =$

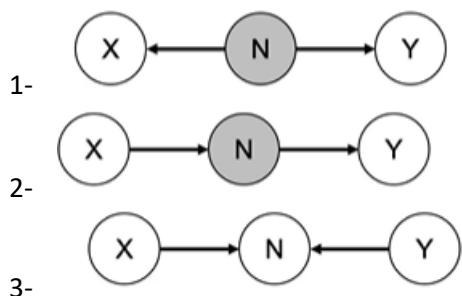
$= P(\text{bad mood} | \text{ticket, red}) P(\text{ticket} | \text{red}) P(\text{red}) =$

$= \text{BOH}$

Reti Bayesiane e D-Separation - ASSIGNMENT

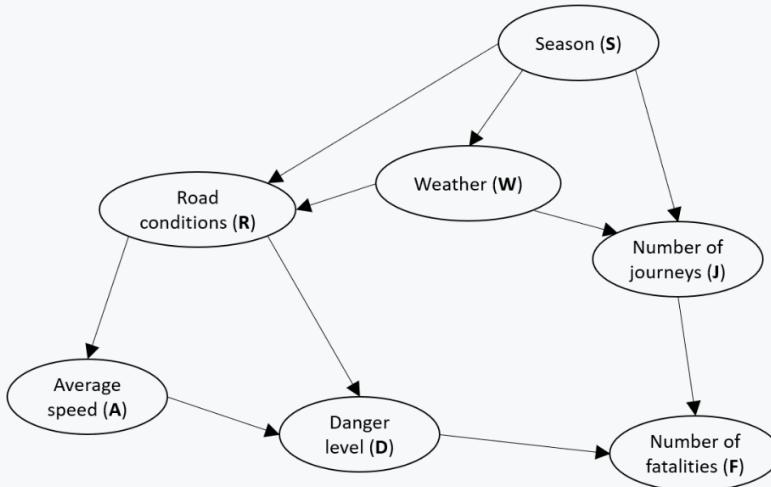
Il concetto di indipendenza condizionale di variabili casuali è collegato al concetto di D-separazione; questa è la proprietà che hanno due variabili di una rete quando queste sono condizionalmente indipendenti. Se due variabili sono condizionalmente indipendenti allora acquisire informazioni su una variabile non mi porta alcun beneficio di conoscenza sull'altra.

X e Z sono d-separate da un insieme E di variabili con evidenza (osservazioni) se e solo se ogni cammino non orientato da X a Z è “bloccato”. Dove un cammino è bloccato se e solo se vale almeno una delle tre seguenti condizioni:



Attenzione che in questo esempio N non è in evidenza

Data la seguente Rete Bayesiana, si risponda (vero o falso) se le seguenti affermazioni di indipendenza condizionale sono riflesse dalla struttura della rete.



R e J d-separano F e S?

- Scegli una risposta:
- Vero ✓
 - Falso

R e J sono in evidenza. I cammini possibili sono:

S,J,F|S,W,J,F|S,W,R,D,F|S,W,R,A,D,F|S,R,D,F|S,R,A,D,F

In ordine i cammini sono D-separati per le seguenti regole:

2|2|2|2|2|2

J d-separa F e W?

- Scegli una risposta:
- Vero
 - Falso ✓

Perché il cammino F,D,R,W non rientra in nessuna delle 3 regole

W e R d-separano D e S?

- Scegli una risposta:
- Vero ✓
 - Falso

Perché tutti i cammini che passano da R e W sono bloccati per la regola 2 e il cammino S|J|F|D è bloccato dalla regola 3 nel punto J|F|D

R d-separa D e J?

- Scegli una risposta:
- Vero ✓
 - Falso

D|F|J per la regola 3, e tutto quello che passa per R per la regola 2

Si indichino tutte le coppie di nodi che sono indipendenti l'uno dall'altro dati i nodi S, R e D.

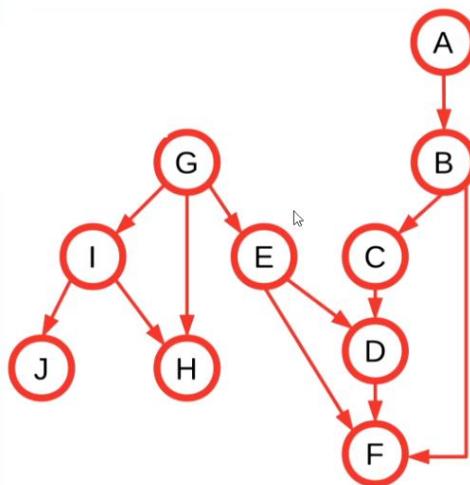
Scegli un'alternativa:

- a. (W,A), (A,F), (J,A) ✓
- b. (W,A), (W,F), (A,F), (J,A)
- c. (W,A), (W,F), (W,J), (A,F), (J,A), (J,F)
- d. (A,F), (J,A)
- e. Insieme vuoto

Se l'esercizio chiedesse di identificare quali nodi siano indipendenti dagli altri allora dobbiamo vedere secondo le tre regole di D-separazione per ogni nodo se esiste almeno una regola verificata e individuare quali sono i nodi che vengono D-separati in questo modo. Quindi in pratica devi fare lo stesso di prima con magari più nodi in evidenza e per ogni coppia della rete.

Domanda 5
Risposta salvata
Punteggio max.:
5,00
Contrassegna domanda

Data la seguente rete bayesiana a variabili booleane:



Si indichi vero/falso per ciascuna delle seguenti affermazioni, motivando la risposta:

- a) F è indipendente da A: Falso
- b) C è indipendente da F: Falso
- c) G è indipendente da A, dato C e F: Falso
- d) Il numero di parametri che caratterizza la rete è 512: Falso

Uguale a prima. La d secondo me è 2^n ma non son sicuro.

Esercitazione Reti Bayesiane, distribuzione di probabilità congiunta e indipendenza

Esercizio 1

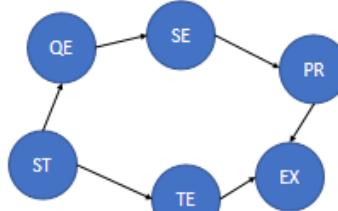
- Sviluppare una rete di Bayes, per calcolare la probabilità che uno studente superi l'esame di MPD. Le proprietà di interesse del problema sono:
 - Il superamento dell'esame ($EX \in \{\text{true}, \text{false}\}$)
 - L'acquisizione di buone capacità pratiche in MPD da parte dello studente ($PR \in \{\text{true}, \text{false}\}$)
 - L'acquisizione di buone capacità teoriche in MPD da parte dello studente ($TE \in \{\text{true}, \text{false}\}$)
 - Lo studente è efficientemente studioso ($ST \in \{\text{true}, \text{false}\}$)
 - La quantità di esercitazioni seguite dallo studente ($QE \in \{\text{molte}, \text{poche}, \text{nessuna}\}$)
 - L'aver fatto un numero sufficiente di esercizi ($SE \in \{\text{true}, \text{false}\}$)
- Costruire una rete bayesiana che rappresenti la conoscenza probabilistica relativa al dominio descritto dalle seguenti relazioni di dipendenza tra le variabili casuali:

- Il superamento dell'esame dipende dalle capacità teoriche e pratiche dello studente.
- Se uno studente è studioso ha buone probabilità di acquisire capacità teoriche.
- Il numero di esercitazioni seguite dipende da quanto uno studente è studioso
- L'aver fatto sufficienti esercizi dipende dal numero di esercitazioni seguite, ed influenza le capacità pratiche dello studente.

Il superamento dell'esame ($EX \in \{\text{true}, \text{false}\}$)
L'acquisizione di buone capacità pratiche in MPD da parte dello studente ($PR \in \{\text{true}, \text{false}\}$)
L'acquisizione di buone capacità teoriche in MPD da parte dello studente ($TE \in \{\text{true}, \text{false}\}$)
Lo studente è efficientemente studioso ($ST \in \{\text{true}, \text{false}\}$)
La quantità di esercitazioni seguite dallo studente ($QE \in \{\text{molte}, \text{poche}, \text{nessuna}\}$)
L'aver fatto un numero sufficiente di esercizi ($SE \in \{\text{true}, \text{false}\}$)

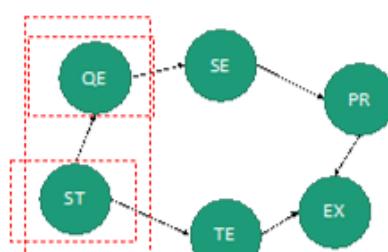
- Costruire una rete bayesiana che rappresenti la conoscenza probabilistica relativa al dominio descritto dalle seguenti relazioni di dipendenza tra le variabili casuali:

- Il superamento dell'esame dipende dalle capacità teoriche e pratiche dello studente.
- Se uno studente è studioso ha buone probabilità di acquisire capacità teoriche.
- Il numero di esercitazioni seguite dipende da quanto uno studente è studioso
- L'aver fatto sufficienti esercizi dipende dal numero di



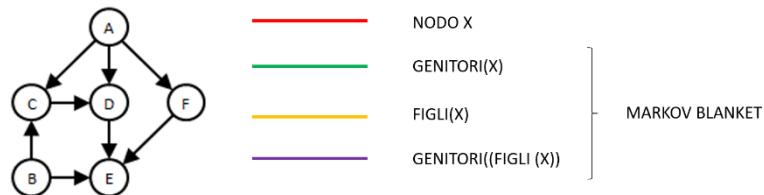
- Scrivere la tabella di probabilità associata al nodo QE, facendo delle ipotesi a piacere sul dominio.

QE = molte	QE = poche	QE = nessuna	ST true false
0.5	0.3	0.2	0.9
0	0.1	0.9	



Facile, basta trovare le dipendenze e fare archi orientati. Per la seconda parte basta vedere i valori che possono assumere le variabili e fare la CPT.

- Data la seguente rete bayesiana, determinare la Markov blanket per ogni nodo.



La Markov Blanket è definita dai genitori del nodo, dai figli del nodo e dai genitori dei figli del nodo.

Esercizio 2

- Data la seguente rete bayesiana, determinare la Markov blanket per ogni nodo.

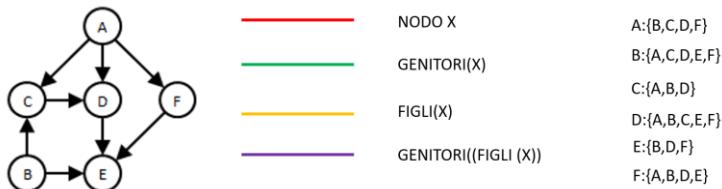


Esercizio 2

- Data la seguente rete bayesiana, determinare la Markov blanket per ogni nodo.



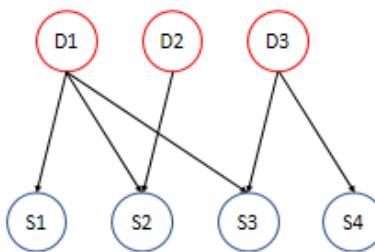
- Data la seguente rete bayesiana, determinare la Markov blanket per ogni nodo.



Ricorda che se si è a conoscenza dello stato (ovvero è in evidenza) della Markov Blanket di un nodo, questo si può dire essere condizionalmente indipendente da ogni altro nodo della rete che non fa parte della Markov Blanket.

- Un paziente si reca dal dottore per sottoporre una patologia, il dottore sospetta 3 possibili malattie come causa della condizione patologica. Le 3 malattie sono D_1 , D_2 , e D_3 , le quali sono marginalmente indipendenti tra loro. Ci sono 4 sintomi S_1 , S_2 , S_3 , e S_4 di cui il dottore vuole verificare la presenza in modo da trovare la causa più probabile per la condizione patologica. I sintomi sono condizionalmente dipendenti alle 3 malattie come segue: S_1 dipende solamente da D_1 , S_2 dipende da D_1 e da D_2 , S_3 dipende da D_1 e da D_3 , e S_4 dipende solamente da D_3 . Si assume che tutte le variabili casuali siano booleane.
 1. Costruire la struttura topologica alla rete bayesiana per il problema descritto.
 2. Scrivere la distribuzione di probabilità congiunta espressa come prodotto delle probabilità condizionate.
 3. Qual è il numero di parametri indipendenti richiesti per descrivere la distribuzione congiunta?
 4. Assumendo che non ci sia indipendenza condizionale tra le variabili, quanti parametri indipendenti sarebbero dunque richiesti?

- Un paziente si reca dal dottore per sottoporre una patologia, il dottore sospetta 3 possibili malattie come causa della condizione patologica. Le 3 malattie sono D_1 , D_2 , e D_3 , le quali sono marginalmente indipendenti tra loro. Ci sono 4 sintomi S_1 , S_2 , S_3 , e S_4 di cui il dottore vuole verificare la presenza in modo da trovare la causa più probabile per la condizione patologica. I sintomi sono condizionalmente dipendenti alle 3 malattie come segue: S_1 dipende solamente da D_1 , S_2 dipende da D_1 e da D_2 , S_3 dipende da D_1 e da D_3 , e S_4 dipende solamente da D_3 . Si assuma che tutte le variabili casuali siano booleane.

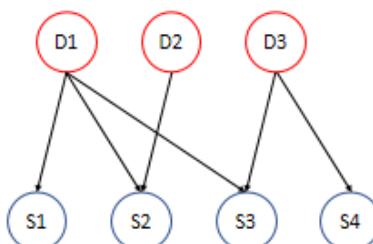


Qua si usa la regola della probabilità congiunta: $P(X,Y) = P(X|Y)*P(Y)$.

- Scrivere la distribuzione di probabilità congiunta espressa come prodotto delle probabilità condizionate.

$$P(D1, D2, D3, S1, S2, S3, S4) =$$

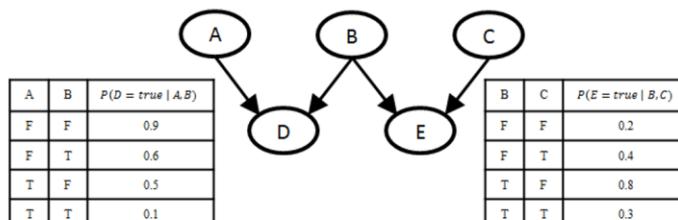
$$= P(D1) * P(D2) * P(D3) * P(S1|D1) * P(S2|D1,D2) * \\ * P(S3|D1,D3) * P(S4|D3)$$



Esercizio 5

- Si consideri la seguente la rete bayesiana in cui tutte le variabili sono booleane.
 - Qual è la probabilità che tutte le variabili siano contemporaneamente false?
 - Qual è la probabilità di A, avendo la conoscenza che tutte le altre variabili sono vere?

$$P(A=\text{true}) = 0.2 \quad P(B = \text{true}) = 0.5 \quad P(C = \text{true}) = 0.8$$



- Qual è la probabilità che tutte le variabili siano contemporaneamente false?

$$P(\neg A, \neg B, \neg C, \neg D, \neg E)$$

$$= P(\neg A)P(\neg B)P(\neg C)P(\neg D|\neg A, \neg B)P(\neg E|\neg B, \neg C)$$

$$= (0.8)(0.5)(0.2)(0.1)(0.8)$$

$$= 0.0064$$

Usando le CPT sostituisci tutti i valori e poi li **moltiplichi** tra di loro.

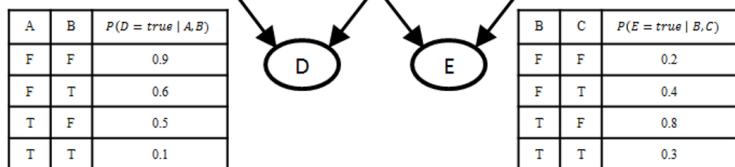
- Qual è la probabilità di A, avendo la conoscenza che tutte le altre variabili sono vere?

$$\begin{aligned}
 & P(\neg A|B, C, D, E) \\
 &= \alpha P(\neg A)P(B)P(C)P(D|\neg A, B)P(E|B, C) \\
 &= \alpha(0.8)(0.5)(0.8)(0.6)(0.3) \\
 &= \underline{\alpha}(0.05760)
 \end{aligned}$$

$$\begin{aligned}
 P(A|B, C, D, E) &= \alpha P(A)P(B)P(C)P(D|A, B)P(E|B, C) \\
 &= \alpha(0.2)(0.5)(0.8)(0.1)(0.3) \\
 &\equiv \alpha(0.00240)
 \end{aligned}$$

$$\alpha = \frac{1}{0.05760 + 0.00240} = \frac{1}{0.06} = 16.66667$$

$< 0.96; 0.04 >$

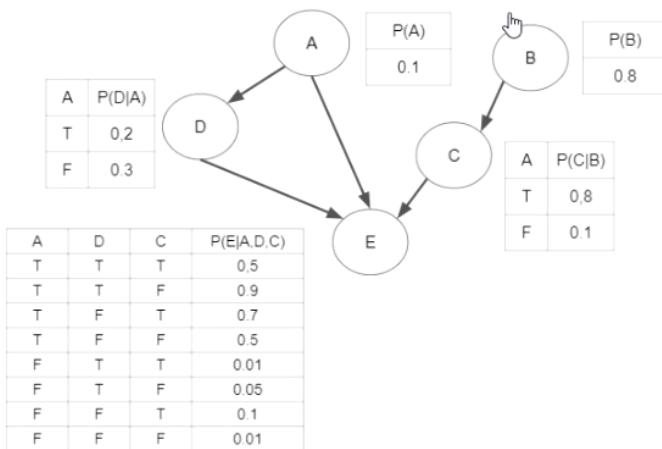


Prima usi la formula della probabilità congiunta per scomporre la probabilità totale, poi anche qui usando le CPT estrai i valori richiesti. Metti **alfa** davanti e fai la versione che tiene conto di A vero e la versione di A falso. Calcoli alfa come **1/ la somma dei due valori ottenuti di A vero e A falso**. Moltiplicherai i risultati per alfa e trovi i due valori.

In realtà alfa puoi evitare di usarlo basta fare A vero / A vero + A falso per il primo membro e A falso / A vero + A falso per il secondo membro.

Inferenza Esatta e Approssimata - ASSIGNMENT

Si consideri la seguente rete Bayesiana. Si risolvano gli esercizi seguenti utilizzando il metodo di inferenza esatta per enumerazione.



- a) Calcolare la probabilità che E è vero sapendo che D=F e B=F.
 b) Calcolare la distribuzione di probabilità congiunta $P(A=F, B=T, C=T, D=T, E=T)$.

Si riporti lo svolgimento dell'esercizio sul foglio protocollo e si riportino le risposte alla domanda a) e b) nel form sottostante.



Per il primo devi guardare la rete di Bayes e scrivere la probabilità come il prodotto di probabilità condizionate. Vediamo come **E** dipenda direttamente da **A,D** e **C**, quindi dobbiamo calcolare la probabilità condizionata $P(E|A,D,C)$ ma sappiamo che **D** è sempre **F** quindi lo lasciamo sempre a **F** e gli altri due no, perciò dobbiamo considerare sia quando **A** è vero sia quando **A** è falso e lo stesso per **C**. Ricordarsi di

mantenere sempre B falso quando si va a calcolare C. Ricordarsi di sommare tra di loro i prodotti delle varie combinazioni possibili. Ovviamente dobbiamo calcolare sia la probabilità che E sia vero dato A,D,C sia E falso dato A,D,C. Poi mettiamo alfa, lo calcoliamo e lo sostituiamo per ottenere la coppia richiesta.

Esercizio

$$P(E | D, \neg B) \rightsquigarrow \alpha$$

$$\alpha \sum_{\substack{A \in \{\text{Vero}, \neg\} \\ D \in \{\text{Vero}, \neg\} \\ C \in \{\text{Vero}, \neg\}}} P(E | A = \text{Vero}, \neg D, \neg B, C = \text{Vero}) +$$

$$= P(E | A = \text{Vero}, \neg D, \neg C) \cdot P(\neg D | A) \cdot P(A) \cdot P(C | \neg B) \cdot P(\neg B) +$$

$$+ P(E | A = \text{Vero}, \neg D, C) \cdot P(\neg D | A) \cdot P(\neg A) \cdot P(C | \neg B) \cdot P(\neg B) +$$

$$+ P(E | A = \neg\text{Vero}, \neg D, \neg C) \cdot P(\neg D | A) \cdot P(A) \cdot P(\neg C | \neg B) \cdot P(\neg B) +$$

$$+ P(E | \neg A, \neg D, \neg C) \cdot P(\neg D | A) \cdot P(\neg A) \cdot P(\neg C | \neg B) \cdot P(\neg B) =$$

$$= 0.7 \cdot 0.8 \cdot 0.1 \cdot 0.1 \cdot 0.2 +$$

$$+ 0.1 \cdot 0.7 \cdot 0.9 \cdot 0.1 \cdot 0.2 +$$

$$+ 0.5 \cdot 0.8 \cdot 0.1 \cdot 0.9 \cdot 0.2 +$$

$$+ 0.01 \cdot 0.7 \cdot 0.9 \cdot 0.9 \cdot 0.2 = 0.00112 + 0.00126 + 0.0072 + 0.001134 =$$

$$= 0.020214 - \alpha$$

$$P(\neg E | A = \text{Vero}, \neg D, \neg B, C = \text{Vero}) = 0.3 \cdot 0.8 \cdot 0.1 \cdot 0.1 \cdot 0.2 +$$

$$0.1 \cdot 0.7 \cdot 0.9 \cdot 0.1 \cdot 0.2 +$$

$$0.5 \cdot 0.8 \cdot 0.1 \cdot 0.9 \cdot 0.2 +$$

$$0.99 \cdot 0.7 \cdot 0.9 \cdot 0.9 \cdot 0.2 = 0.00048 + 0.00126 + 0.0072 +$$

$$+ 0.122266 = 0.131206 \rightsquigarrow \alpha$$

Come spiegato prima $P(E | \neg D, \neg B)$ per essere calcolato bisogna valorizzare sempre a falsi D e B e poi fare il calcolo della probabilità di E condizionata dai suoi genitori, ovvero A,C e D. Dato che non abbiamo fissato A e C allora nel calcolo dobbiamo considerare le loro possibili variazioni combinate tra loro. Nella prima riga abbiamo A vero e C vero, nella seconda A falso e C vero e così via... Quindi sommiamo tra di loro le combinazioni per ottenere il caso in cui E è vero. Facciamo lo stesso per il caso in cui E falso. Nei risultati ottenuti mettiamo alfa davanti. Calcoliamo alfa. Sostituiamo alfa e vediamo la coppia di risultati.

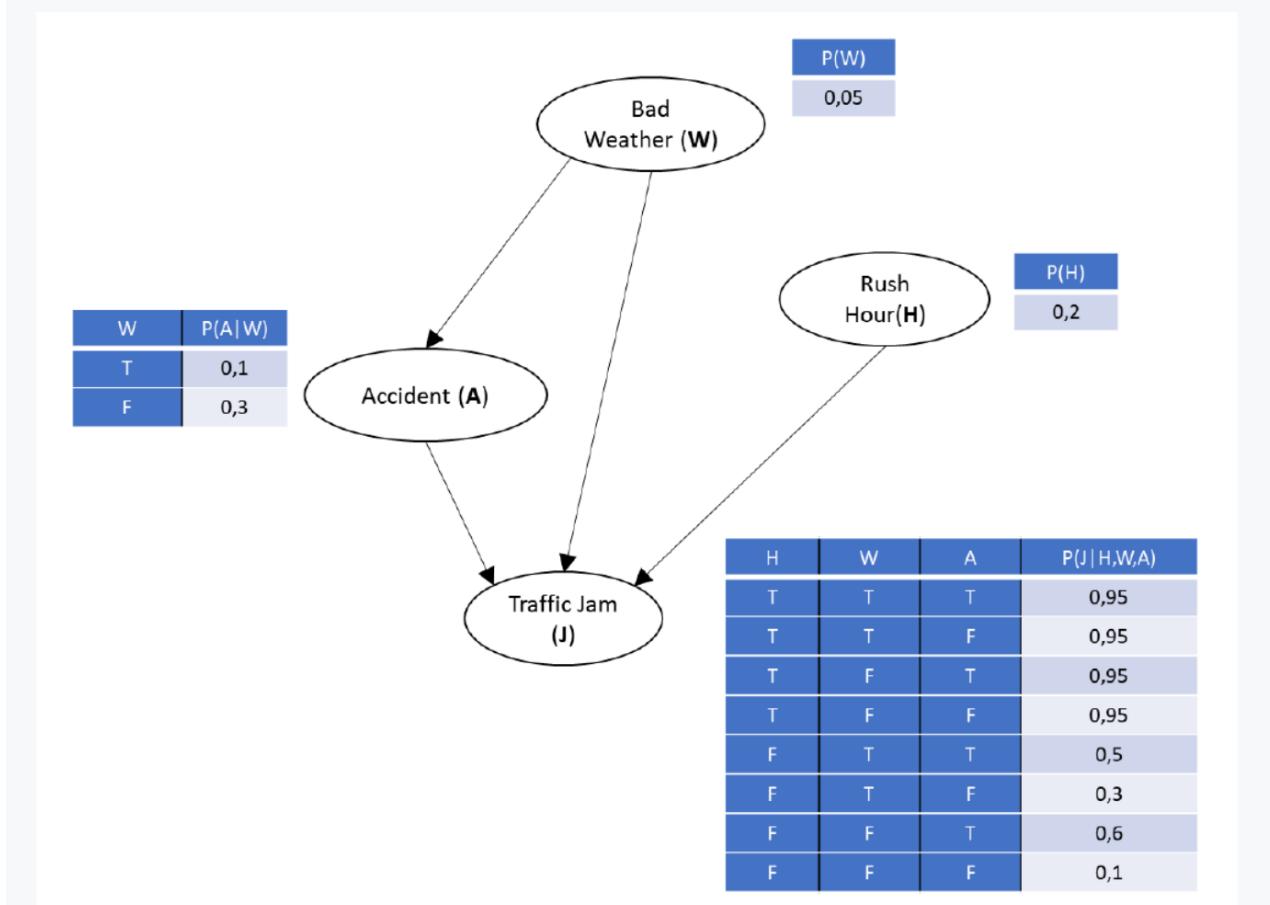
$$\rightsquigarrow \alpha = \frac{1}{0.020214 + 0.131206} = 7.58035173$$

$$\rightsquigarrow \langle 0.081, 0.919 \rangle$$

C'è un errore nella parte con $\neg E$, seconda riga di prodotti, messo 0.1 a $P(\neg E | \neg A, \neg D, C)$, dovrebbe essere 0.9. $0.131206 \rightarrow \alpha = 7.04 \rightarrow 0.07545$ il risultato finale.

Questo esercizio non faceva parte dell'assignment di quest'anno. In ogni caso devi procedere come per l'esercizio che viene qui sotto. Tieni conto di tutte le possibili combinazioni delle variabili non in evidenza e fai i conti. Non so perché usa alfa qua e dopo no.

Si consideri la seguente rete bayesiana:



Si calcoli la probabilità di rimanere bloccati nel traffico ($J=True$) data l'evidenza che è l'ora di punta ($H=True$), c'è un incidente ($A=True$) e c'è brutto tempo ($W=True$).

◀

Scegli un'alternativa:

- a. 0,5
- b. 0,95 ✓
- c. 0,877
- d. 1
- e. 0,05

Risposta corretta.

Per rispondere correttamente a questo quesito, occorre guardare la CPT di J . In particolare, occorre trovare la cella corrispondente a $P(J=true | H=true, A=true, W=true)$.

La risposta corretta è: 0,95

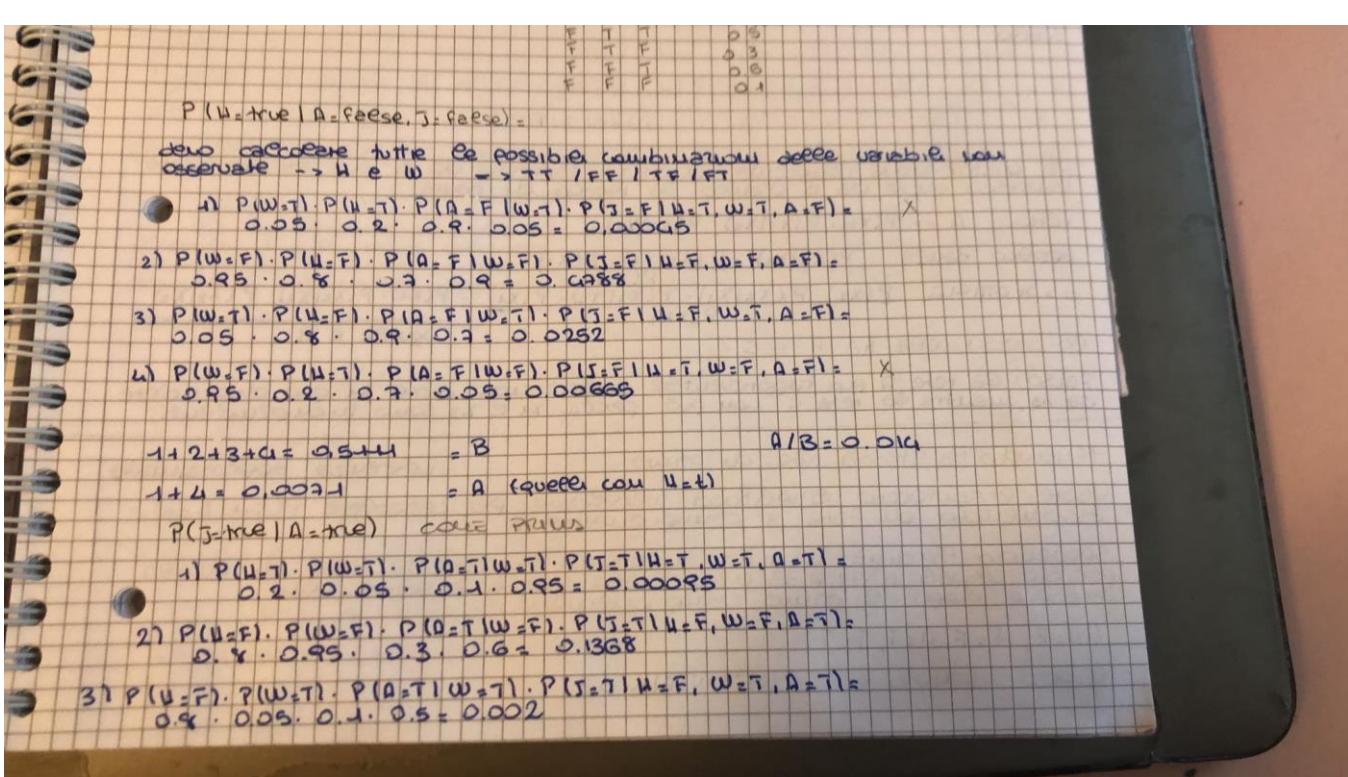
Si calcoli il valore di $P(H=\text{true} | A=\text{False}, J=\text{False})$.



Scegli un'alternativa:

- a. 0,00665
- b. 0,5111
- c. 0,0071
- d. 0,504
- e. 0,014 ✓

In questo caso dobbiamo prendere in considerazione tutte le possibili combinazioni delle variabili di cui non abbiamo un valore fisso, ovvero H e W. Quindi dobbiamo calcolare tutte le combinazioni possibili per H vero, H falso e W vero, W falso. Avremo sempre quindi le due probabilità di cui non abbiamo valore fisso all'inizio che variano in ognuno dei 4 casi di combinazione possibile e poi avremo A e J il cui valore sarà sempre fisso che però sono condizionate dai valori che variano di H e W. Una volta calcolate le 4 combinazioni le sommiamo per avere il totale e, dato che è richiesto di calcolare H true andiamo a sommare le due combinazioni di H true e andiamo a dividere il risultato per il totale della somma delle 4 combinazioni.



Si calcoli la probabilità di rimanere bloccato nel traffico quando c'è un incidente sulla strada.



Scegli un'alternativa:

- a. 0,67 ✓
- b. 0,57
- c. 0,95
- d. 0,42
- e. 0,5

Le prime 3 combinazioni di questo esercizio sono mostrate nell'immagine sopra, il resto nell'immagine sotto.

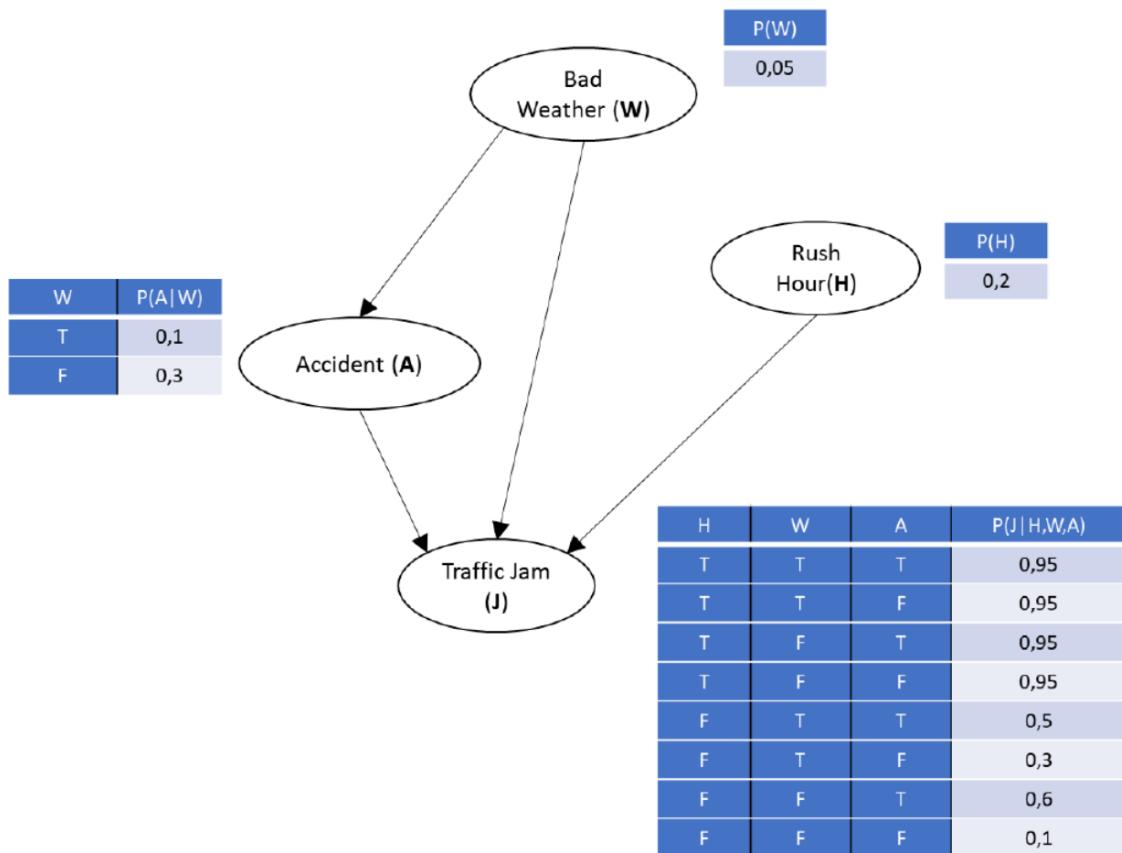
4) $P(H=T) \cdot P(W=F) \cdot P(A=\bar{T} | W=F) \cdot P(J=T | H=\bar{T}, W=F, A=\bar{T})$
 $0.2 \cdot 0.95 \cdot 0.3 \cdot 0.95 = 0.05445$
 $1+2+3+4 = 0.1939$
 Ora calcoliamo tutti i casi con $J=F$
 1.1) $0.2 \cdot 0.05 \cdot 0.1 \cdot 0.05 = 0.00005$
 2.1) $0.8 \cdot 0.95 \cdot 0.3 \cdot 0.0 = 0.0912$
 3.1) $0.8 \cdot 0.05 \cdot 0.1 \cdot 0.5 = 0.002$
 4.1) $0.2 \cdot 0.95 \cdot 0.3 \cdot 0.05 = 0.00285$
 $1.1 + 2.1 + 3.1 + 4.1 = 0.0961$
 $0.1939 / 0.1939 + 0.0961 = 0.67$

Anche in questo caso dobbiamo prendere in considerazione le variabili non osservate, ovvero quelle che non stanno dopo la | che sono H, W e J. Qui ovviamente ci saranno molte più combinazioni da calcolare e le possiamo dividere in due parti: una che identifica J true e una che identifica J false. Questa scelta viene fatta perché alla fine dovremo sommare le combinazioni per J true e J false e fare la divisione solita di J true/Jtot.

Vediamo come nelle prime 4 combinazioni calcoliamo per J true le combinazioni al variare di W e H. Allo stesso modo nelle 4 combinazioni finali (solo coi numeri) calcoliamo per J false tutte le combinazioni al variare di W e H.

Una volta fatto questo basta sommare J true e J false per avere J tot e poi, dato che volgiamo la probabilità di J true, andiamo a dividere J true per J tot.

Si consideri la seguente rete Bayesiana.



Utilizzare l'approccio di **likelihood weighting** per calcolare $P(J=True | A=True)$ assumendo di fare 3 campionamenti.

(W, H, A, J) è un ordinamento topologico dei nodi della rete Bayesiana che può essere seguito durante il campionamento. Quali altri ordinamenti sono comunque corretti?

◀

Scegli una o più alternative:

- a. H, J, W, A
- b. H, W, A, J ✓
- c. J, A, H, W
- d. W, A, J, H
- e. W, A, H, J ✓

Per gli ordinamenti topologici basta solo mettere in ordine i nodi assicurandosi di mettere prima i nodi senza genitori e poi i nodi i quali genitori sono già stati elencati.

Primo campionamento.

Seguendo l'ordine topologico **(W, H, A, J)**, si consideri la seguente lista di numeri pseudocasuali: **(0,71, 0,33, 0,24)**

Selezionare il valore corretto associato a ciascuna variabile per il primo campione.

W:	False ↴	✓
H:	False ↴	✓
A:	True ↴	✓
J:	True ↴	✓

Guardando i valori delle CPT genero i grafici mettendo sulla X per primo il valore più alto (a prescindere se sia del vero o del falso) e poi utilizzando i numeri pseudo casuali vedo se ricade nell'intervallo del vero o falso nel grafico. A è true perché dato nella traccia e quindi l'ultimo valore pseudo casuale va associato a J = 0.24

Il peso ottenuto dal primo campionamento è:

Scegli un'alternativa:

- a. 0.1
- b. 0.3 ✓
- c. 0.05
- d. 0.7
- e. 0.9

Per calcolare il peso devo usare SEMPRE le variabili osservate, in questo caso A, e calcolare il loro valore che resta fisso sulla base dei valori delle variabili genitori ottenute dai numeri pseudo casuali, in questo caso W false. Si fa la formula della probabilità condizionata $P(A=t | W=f)$.

Utilizziamo l'approccio di **MCMC** per calcolare $P(J=True | A=True)$. Quali variabili verranno iterativamente campionate?

Scegli un'alternativa:

- a. J, A
- b. W, H, J ✓
- c. W, H, A
- d. W, H, J, A
- e. W, H

Le variabili che vengono iterativamente campionate sono quelle che non sono in evidenza W e H, compresa quella risultato ovvero J.

Si assume il seguente stato iniziale $S_0 = [W=F, H=F, A=T, J=T]$.

Estraiamo il numero 0.74 e campioniamo W utilizzando l'approccio di MCMC. Allora lo stato successivo S_1 sarà:

W:	False	✓
H:	False	✓
A:	True	✓
J:	True	✓

Qui stiamo campionando W quindi calcoliamo i due casi in cui W true e W false, mantenendo fissi i valori delle altre variabili definite dallo stato iniziale e che fanno parte della Markov Blanket di W (**forse sono solo i genitori e i figli della variabile campionata**) , $P(w=f | h=f, a=t, j=t)$ e $P(w=t | h=f, a=t, j=t)$. Calcolate queste due utilizzando i valori della CPT le sommo per trovare il TOT e poi faccio W true/TOT e W false/TOT per

normalizzare i valori. Una volta fatto tutto uso il numero pseudo casuale per vedere dove ricade nel grafico definito dai due nuovi numeri trovati di W.

Si consideri la rete bayesiana dell'esercizio precedente.

Stimare $P(J=True|A=True)$ tramite il metodo di rejection sampling.

Seguendo l'ordine topologico (W, H, A, J), si consideri la seguente lista ordinata di numero pseudocasuali per generare un campione: [0.71, 0.33, 0.24, 0.01].

Il campione generato verrà rigettato.

Scegli una risposta:

- Vero ✓
 Falso

Associamo i numeri pseudo casuali alle variabili in ordine topologico. Facciamo il grafico per ognuna delle variabili e vediamo dove ricade il numero pseudo casuale per assegnare il valore di vero o falso. Vediamo se il valore ottenuto con il numero pseudo casuale di A coincide con quello delle ipotesi iniziali, ovvero $A=True$. In questo caso non coincide quindi viene rigettato.

Si assuma di aver generato 1000 campioni così suddivisi:

- 430 campioni con $J = T$ e $A = True$
- 320 campioni con $J = T$ e $A = False$
- 210 campioni con $J = F$ e $A = True$
- 40 campioni con $J = F$ e $A = False$

Qual è il valore di $P(J=True | A = True)$?

◀

Scegli un'alternativa:

- a. 0.67 ✓
 b. 0.43
 c. 0.79
 d. 0.49
 e. 0.33

Per calcolare il valore di $P(J=True | A=True)$, possiamo utilizzare i campioni generati e applicare la formula di probabilità condizionata. $P(J=True | A=True) = (\text{Numero di campioni con } J=True \text{ e } A=True) / (\text{Numero di campioni con } A=True)$

Dai dati forniti, abbiamo:

- Numero di campioni con $J=True$ e $A=True$ = 430
- Numero di campioni con $A=True$ = $430 + 210 = 640$

Applicando la formula, otteniamo:

$$P(J=True \mid A=True) = 430 / 640 = 0.67$$

Sempre assumendo di aver generato 1000 campioni così suddivisi:

- 430 campioni con J = T e A = True
- 320 campioni con J = T e A = False
- 210 campioni con J = F e A = True
- 40 campioni con J = F e A = False

Qual è invece il valore di $P(J=True)$ senza avere evidenze?



Scegli un'alternativa:

- a. 0.43
- b. 0.25
- c. 0.75 ✓
- d. 0.49
- e. 0.67

Per calcolare il valore di $P(J=True)$ senza avere evidenze, possiamo utilizzare i campioni generati e applicare la formula di probabilità marginale. La formula è definita come:

$$P(J=True) = (\text{Numero di campioni con } J=\text{True}) / (\text{Numero totale di campioni})$$

Dai dati forniti, abbiamo:

- Numero di campioni con $J=True = 430 + 320 = 750$
- Numero totale di campioni = 1000

Applicando la formula, otteniamo:

$$P(J=True) = 750 / 1000 = 0.75$$

Esercitazione Inferenza Esatta e Approssimata

Esercizio 1

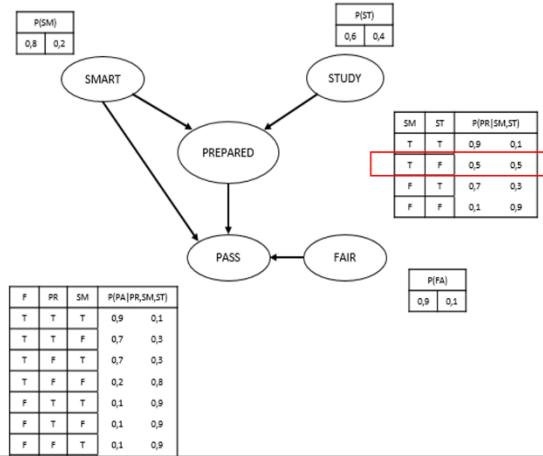
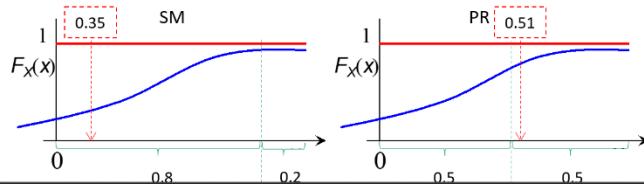
- Data la seguente rete bayesiana, generare N=1000 campioni mediante Direct Sampling.

ORDINE TOPOLOGICO: SM, ST, PR, FA, PA

RANDOM NUMBER GENERATOR:

- [0.35, 0.76, 0.51, 0.44, 0.08] → S1

S1 → [0.35 → SM = T ; 0.76 → ST = F ; 0.51 → PR = F ; 0.44 → FA = T ; 0.08 → PA = T]



Per prima cosa si definisce l'ordine topologico ovvero si mettono in ordine i nodi della rete mettendo per primi quelli che non hanno genitori e poi i loro figli e così via. Una volta definito l'ordine assegniamo ai nodi ordinati i valori dei numeri pseudo casuali.

Successivamente creiamo il grafico con l'onda. Sulla X mettiamo i valori della CPT del nodo che stiamo prendendo in considerazione, prestando attenzione a mettere il valore più grande per primo (a prescindere che sia true o false). Vedo poi dove cade il numero pseudo casuale rispetto al grafico appena creato e assegno true o false in base a questo. Il valore true o false andrà a costituire il primo elemento del nostro sampling. Allo stesso modo per ogni nodo della rete faccio la stessa cosa e vedo cosa mi esce alla fine come combinazione di true e false e vedo qual è il valore del nodo finale. Ricordarsi che nel caso una CPT sia composta da più elementi di usare la riga che corrisponde ai valori T o F trovati nei passaggi precedenti dei genitori.

Esercizio 1

- Data la seguente rete bayesiana, generare N=1000 campioni mediante Direct Sampling.

ORDINE TOPOLOGICO: SM, ST, PR, FA, PA

RANDOM NUMBER GENERATOR:

- [0.35, 0.76, 0.51, 0.44, 0.08] → S1

S1 → [0.35 → SM = T ; 0.76 → ST = F ; 0.51 → PR = F ; 0.44 → FA = T ; 0.08 → PA = T]

- [0.28, 0.03, 0.92, 0.92, 0.42] → S2

S2 → [0.28 → SM = T ; 0.03 → ST = T ; 0.92 → PR = F ; 0.92 → FA = F ; 0.42 → PA = F]

...

S1000

Con 1000 campioni:

- 690 PA = T
- 310 PA = F

< 0.69 ; 0.31 >

Esercizio 2

- Data la seguente rete bayesiana, calcolare $P(PA = T | ST = F)$ tramite Rejection Sampling.

ORDINE TOPOLOGICO: SM, ST, PR, FA, PA

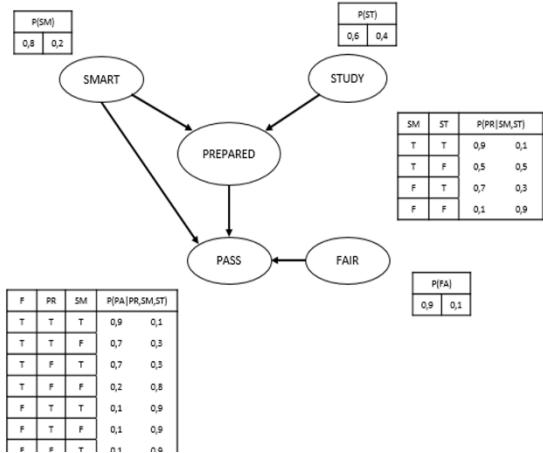
RANDOM NUMBER GENERATOR:

- $[0.35, 0.76, 0.51, 0.44, 0.08] \rightarrow S1$

$S1 \rightarrow [0.35 \rightarrow SM = T ; 0.76 \rightarrow ST = F ; 0.51 \rightarrow PR = F ; 0.44 \rightarrow FA = T ; 0.08 \rightarrow PA = T] \rightarrow \text{ACCEPT}$

- $[0.28, 0.03, 0.92, 0.92, 0.42] \rightarrow S2$

$S2 \rightarrow [0.28 \rightarrow SM = T ; 0.03 \rightarrow ST = T ; 0.92 \rightarrow PR = F ; 0.92 \rightarrow FA = F ; 0.42 \rightarrow PA = F] \rightarrow \text{REJECT}$



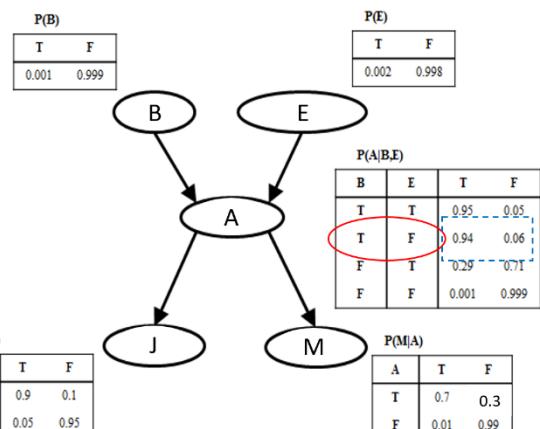
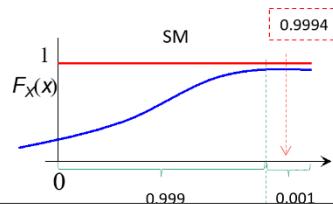
Allo stesso modo di prima creiamo l'ordine topologico e poi, sulla base dei numeri pseudo casuali, andiamo ad attribuire i valori ai nodi e vediamo se questi sono veri o falsi. In questo caso però essendo rejection sampling abbiamo che ST è sempre Falso, pertanto accetteremo tutti quei sample per cui ST è falso e rifiuteremo tutti quei sample per cui ST è vero.

Esercizio 3

- Data la seguente rete bayesiana, calcolare $P(B = \text{true} | J = \text{true}, M = \text{false})$ tramite Likelihood Weighting.

Ordine topologico: B, E, A, J, M

- Fixo le evidenze
- Genero 1000 campioni in accordo alle distribuzioni di probabilità (usando i numeri pseudocasuali) e seguendo l'ordine topologico:
 - $S1 = [0.9994; 0.89; 0.3; /; /] \rightarrow [\text{true}, \text{false}, \text{true}; \text{true}, \text{false}]$



Anche qui per i Likelihood Wighting, devo iniziare alla stessa maniera degli esercizi precedenti. Genero un ordine topologico poi, sulla base dei numeri pseudo casuali che in questo caso tengono contro delle evidenze, faccio il solito grafico mettendo sulla X per primo il valore più grande, e vedo il valore dei nodi associati a quel numero pseudo casuale.

- Data la seguente rete bayesiana, calcolare $P(B = \text{true} | J = \text{true}, M = \text{false})$ tramite Likelihood Weighting.

Ordine topologico: B, E, A, J, M

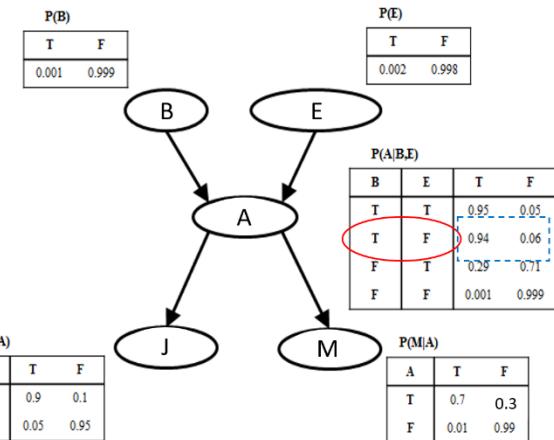
1. Fisso le evidenze

2. Genero 1000 campioni in accordo alle distribuzioni di probabilità (usando i numeri pseudocasuali) e seguendo l'ordine topologico:

- $S_1 = [0.9994; 0.89; 0.3; /; /] \rightarrow [\text{true}, \text{false}, \text{true}, \text{true}, \text{false}]$
- $w_{s1} = P(J=\text{true} | A=\text{true}) * P(M=\text{false} | A=\text{true}) = 0.9 * 0.3 = 0.27$

- $S_2 = [0.7; 0.9; 0.6; /; /] \rightarrow [\text{false}, \text{true}, \text{false}, \text{true}, \text{false}]$

$$w_{s2} = P(J=\text{true} | A=\text{false}) * P(M=\text{false} | A=\text{false}) = 0.05 * 0.99 = 0.0495$$



Ciò che differenzia il likelihood dai precedenti è che **utilizzo le variabili con evidenza per associare un peso al campione generato e vedere quanto affidabile è questo campione**. Quindi in questo caso **per dare il peso w al nostro campione S1 andiamo a moltiplicare tra di loro tutte le variabili di cui siamo a conoscenza ovvero J e M la cui probabilità è influenzata dalla variabile il cui valore è stato definito dai numeri pseudo casuali. I valori da moltiplicare ovviamente li tiriamo fuori dalla CPT**.

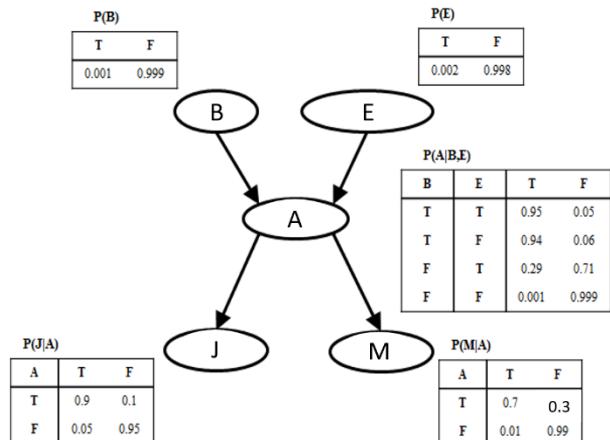
Esercizio 3

- Data la seguente rete bayesiana, calcolare $P(B = \text{true} | J = \text{true}, M = \text{false})$ tramite Likelihood Weighting.

Stimiamo quindi $P(B=\text{true} | J=\text{true}, M=\text{false})$:

$$P(B = \text{true} | J = \text{true}, M = \text{false}) = \frac{\sum_{S_i: B=\text{true}} w_{S_i}}{\sum_{S_i} w_{S_i}}$$

$$= \frac{0.27}{0.27 + 0.0495} = 0.8450$$

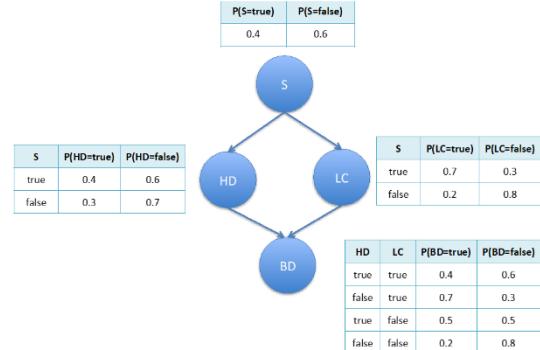


Per concludere l'esercizio la probabilità richiesta da calcolare è definita come la divisione tra i pesi di tutti i campioni che hanno B true fratto il peso di tutti i campioni in generale.

Esercizio 4

- Data la seguente rete bayesiana, calcolare $P(HD|S=t, BD=t)$ mediante il metodo Markov Chain Monte Carlo.
- Fisso le evidenze e campiono una variabile alla volta.
 - Inizializzo in modo random le variabili non osservate, $LC=t$ e $HD=f$, e determino quindi lo **stato iniziale** $S_0 = [S=t, HD=f, LC=t, BD=t]$
 - Campiono le variabili non osservate (variabili nascoste e la variabile query) dato la relativa Markov Blanket

$$P(x|MB(x)) = \alpha \cdot P(x|Pa(x)) \cdot \prod_{y \in Children(x)} P(y|Pa(y))$$



Notare che ci viene chiesta la probabilità di HD senza specificare se true o false. Per MCMC non si deve fare ordinamento topologico. In questo tipo di approccio fisso analogamente agli esempi precedenti le variabili con evidenza ovvero S e BD in questo caso, poi inizializzo in modo random le variabili non osservate ovvero HD e LC. Inizializzare non vuol dire usare numeri pseudo casuali ma letteralmente mettere a true o false a caso le variabili. In questo modo possiamo definire lo stato iniziale S_0 come la composizione delle true e false delle variabili. Ora in maniera iterativa campioniamo le variabili non osservate data la relativa Markov Blanket, quindi LC e HD.

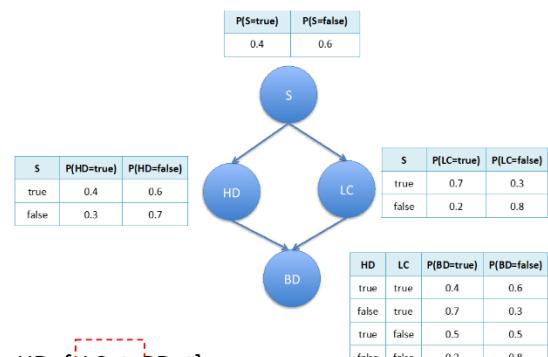
Esercizio 4

- Data la seguente rete bayesiana, calcolare $P(HD|S=t, BD=t)$ mediante il metodo Markov Chain Monte Carlo.

- Campiono LC dato le variabili della sua MB:

$$\begin{aligned}
 & P(LC = t | S = t, HD = f, BD = t) = \\
 &= P(LC = t | S = t)P(BD = t | LC = t, HD = f) \\
 &= 0.7 * 0.7 = 0.49 \\
 & P(LC = f | S = t, HD = f, BD = t) = \\
 &= P(LC = f | S = t)P(BD = t | LC = f, HD = f) \\
 &= 0.3 * 0.2 = 0.06
 \end{aligned}$$

$$\begin{aligned}
 & P(LC | S = t, HD = f, BD = t) = \\
 & \alpha < 0.49; 0.06 > = < 0.89; 0.11 >
 \end{aligned}$$



- campiono LC, ottenendo $LC=f$

$$S_0 = [S=t, HD=f, LC=t, BD=t]$$

$$S_1 = [S=t, HD=f, LC=f, BD=t]$$

Il passaggio iterativo lo iniziamo ad applicare ad LC. Vediamo che LC inizializzato in modo random era true, quindi vediamo qual è la probabilità di LC true e false condizionata alla sua Markov Blanket (che sono i genitori, i figli e i genitori dei figli). Ovviamente i valori delle altre variabili non variano ma vengono utilizzate quelle ottenute dallo step di generazione randomica precedente e dalle variabili che ovviamente devono restare fisse.

Una volta calcolato LC true e false aggiungiamo alfa davanti ai due risultati, calcoliamo alfa come $1/LC\text{ true} + LC\text{ false}$ e sostituiamo il risultato di alfa per ottenere la coppia di valori finali. In realtà puoi non usare alfa ma fare $LC\text{ true} / LC\text{ true} + LC\text{ false}$ per calcolare il primo membro della coppia e $LC\text{ false} / LC\text{ true} + LC\text{ false}$ per il secondo membro.

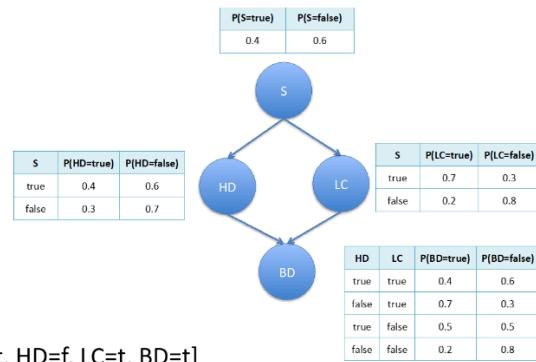
Questo processo ci fa ottenere una nuova distribuzione della variabile LC e quindi devo generare un numero pseudo casuale e vedere dove ricade nella nuova distribuzione per assegnare un nuovo valore di true o false alla variabile LC. Così ottengo lo stato S1.

- Data la seguente rete bayesiana, calcolare $P(HD|S=t, BD=t)$ mediante il metodo Markov Chain Monte Carlo.

4. Campiono HD dato le variabili della sua MB:

$$\begin{aligned}
 & P(HD = t|S = t, LC = f, BD = t) \\
 &= P(HD = t|S = t)P(BD = t|LC = f, HD = t) \\
 &= 0.4 * 0.5 = 0.2 \\
 & P(HD = f|S = t, LC = f, BD = t) \\
 &= P(HD = f|S = t)P(BD = t|LC = f, HD = f) \\
 &= 0.6 * 0.7 = 0.42 \\
 & P(HD|S = t, LC = f, BD = t) = \\
 & \alpha < 0.2; 0.42 > = < 0.32; 0.68 >
 \end{aligned}$$

campiono HD, ottenendo HD=t



$S_0 = [S=t, HD=f, LC=t, BD=t]$

$S_1 = [S=t, HD=f, LC=f, BD=t]$

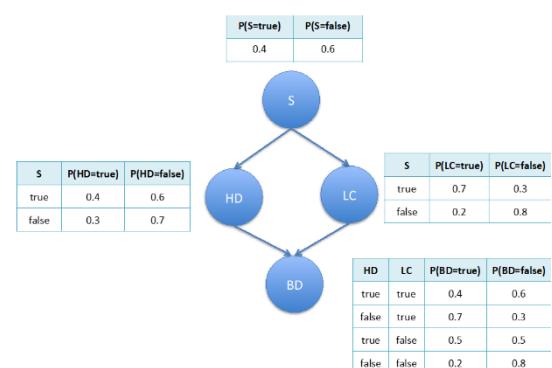
$S_2 = [S=t, HD=t, LC=f, BD=t]$

Ora devo eseguire lo stesso passaggio spiegato prima ma per la variabile HD e applicarla allo stato S1, quindi dovrò utilizzare come valori fissi i valori delle variabili definiti in S1. In questo modo otterrò lo stato S2 e così posso iterare all'infinito.

- Data la seguente rete bayesiana, calcolare $P(HD|S=t, BD=t)$ mediante il metodo Markov Chain Monte Carlo.

5. Itero il processo di campionamento generando $N=1000$ stati della rete. Supponendo che in 800 campioni $HD=t$ e 200 campioni $HD=f$:

$$P(HD|S=t, BD=t) = <0.8; 0.2>$$



Catene di Markov- ASSIGNMENT

Considerate il seguente gioco della “scalata”. Ci sono 5 livelli nel gioco, dove il livello 1 è il più basso (base) mentre il livello 5 è il più alto (cima). Un giocatore parte dalla base. Ad ogni iterazione viene lanciata una moneta non truccata. Se la moneta indica “testa” il giocatore sale di un livello. Se la moneta indica “croce” il giocatore si ritrova di nuovo alla base.

Una volta arrivato alla cima il giocatore torna alla base se dal lancio della moneta esce “croce”, altrimenti rimane dove si trova (sempre in cima).

Trovare la matrice delle probabilità di transizione P.

	1	2	3	4	5
1	1/2	1/2	0	0	0
2	1/2	0	1/2	0	0
3	1/2	0	0	1/2	0
4	1/2	0	0	0	1/2
5	1/2	0	0	0	1/2

Trovare la matrice delle probabilità di transizione a due-step P2

	1	2	3	4	5
1	1/2	1/4	1/4	0	0
2	1/2	1/4	0	1/4	0
3	1/2	1/4	0	0	1/4
4	1/2	1/4	0	0	1/4
5	1/2	1/4	0	0	1/4

Bisogna fare il prodotto tra matrici. Per fare il prodotto si **utilizzano le righe e le colonne delle matrici che vengono moltiplicate**. Scriveremo gli elementi della tabella risultato riga per riga. Gli elementi vengono ottenuti calcolando la somma del prodotto degli elementi di una riga della tabella per gli elementi delle colonne dell'altra tabella. Ad **esempio** per ottenere la **prima riga** della nuova tabella risultato dovremmo **moltiplicare gli elementi della prima riga della prima tabella per gli elementi delle colonne della seconda tabella a scorrere**, nel senso faccio: elementi della prima riga X elementi della prima colonna; elementi della prima riga X elementi della seconda colonna; elementi della prima riga X elementi della terza colonna; e così via fino all'ultima colonna. Arriverò quindi ad avere 5 numeri che sono gli elementi della **prima riga della nuova tabella**. Ricorda che devi sommare le moltiplicazioni fatte tra righe e colonne nel senso: 1 elemento prima riga X 1 elemento prima colonna + 2 elemento prima riga X 2 elemento prima colonna + 3 elemento prima riga X 3 elemento prima colonna e così via. La somma di queste moltiplicazioni darà vita al primo elemento della prima riga della nuova tabella. Iteri tenendo fissa la riga e variando le colonne per ottenere i 5 numeri che servono per popolare la **prima riga** della nuova tabella. **In altre parole fai riga1 X colonna 1, riga 1 X colonna 2, riga1 X colonna 3, riga 1 X colonna 4, riga 1 X colonna 5 e ottieni la prima riga della tabella soluzione. Poi passi alla riga 2 e moltipichi per ogni colonna e così via.**

Trovare la distribuzione stazionaria.

Sia π la distribuzione stazionaria e π_i la probabilità relativa allo stato i , allora

$\pi_1 =$	0.5	▼	✓
$\pi_2 =$	0.25	▼	✓
$\pi_3 =$	0.125	▼	✓
$\pi_4 =$	0.0625	▼	✓
$\pi_5 =$	0.0625	▼	✓

Per trovare la distribuzione stazionaria dobbiamo **fare uso della matrice delle probabilità di base** dell'esercizio precedente. Dobbiamo quindi impostare **un sistema ad N incognite** dove N è il numero di stati **in cui può vertere il sistema, in questo caso 5**. Avremo quindi un sistema dove dobbiamo **trovare il valore di**

$\pi_1, \pi_2, \pi_3, \pi_4, \pi_5$. Il valore di questi è determinato dalla somma delle probabilità di partire da un qualunque stato del sistema e raggiungere lo stato identificato dal π . Abbiamo quindi per lo stato 1 la somma delle probabilità di partire dallo stato 1,2,3,4,5 e arrivare sempre allo stato 1 a cui ci aggiungiamo per ogni stato di partenza un π con il numero dello stato da cui partiamo. Una volta fatto questo per ogni incognita mettiamo sempre alla fine che la somma di tutte le incognite π è =1 per garantire stocasticità, ovvero che la somma delle probabilità di ogni riga è pari ad 1. A questo punto andiamo a sostituire le probabilità di partire da uno stato e raggiungere un altro utilizzando i valori della matrice di probabilità di partenza; in generale uso gli elementi delle colonne che rappresentano lo stato incognita che vogliamo trovare.

Si prende il grafico della prima risposta e si risolve il sistema di equazioni.

$$\pi_1 P_{11} + \pi_2 P_{21} + \pi_3 P_{31} + \pi_4 P_{41} + \pi_5 P_{51} = \pi_1$$

$$\pi_1 P_{12} + \pi_2 P_{22} + \pi_3 P_{32} + \pi_4 P_{42} + \pi_5 P_{52} = \pi_2$$

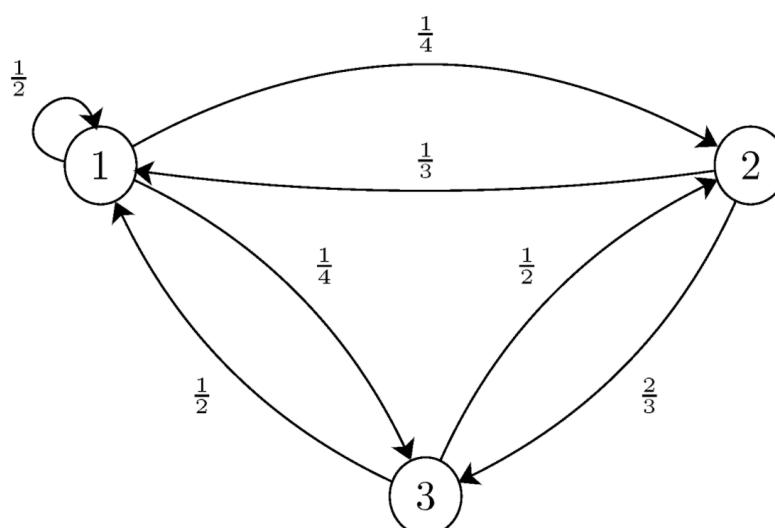
$$\pi_1 P_{13} + \pi_2 P_{23} + \pi_3 P_{33} + \pi_4 P_{43} + \pi_5 P_{53} = \pi_3$$

$$\pi_1 P_{14} + \pi_2 P_{24} + \pi_3 P_{34} + \pi_4 P_{44} + \pi_5 P_{54} = \pi_4$$

$$\pi_1 P_{15} + \pi_2 P_{25} + \pi_3 P_{35} + \pi_4 P_{45} + \pi_5 P_{55} = \pi_5$$

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1$$

Data la seguente rete:



La catena è **irriducibile**? Vero. La catena in figura è una catena irriducibile in quanto tutti i suoi stati sono comunicanti fra loro. In particolare se è possibile raggiungere qualsiasi stato partendo da qualsiasi altro stato in un numero finito di passi allora la catena è irriducibile.

La catena è **periodica**? Falso. La catena in figura non è periodica in quanto lo stato 1 ha un periodo $k=1$, mentre gli stati 2 e 3, a causa del cappio presente nello stato 1, possiede dei cammini il cui MCD è 1 rendendo quindi anche questi aperiodici. Una catena per essere periodica deve essere innanzitutto ergodica, ovvero che sia possibile raggiungere un qualsiasi stato partendo da un qualsiasi altro stato

facendo un numero qualunque di passi intermedi da altri stati e che la catena sia ricorrente positiva ovvero che ritorni a visitare gli stati della catena con una certa frequenza. Se è ergodica bisogna elevare a potenza la matrice di transizione N volte finché i valori al suo interno non si stabilizzano, ovvero restano più o meno uguali. Fatto questo vedo la diagonale principale della matrice e vedo se i numeri sono positivi, se lo sono allora è **aperiodica**. Quindi per avere una catena **periodica** nella diagonale dobbiamo avere degli 0.

Si indichi vero/falso per le seguenti affermazioni, motivandone la risposta:

- 1- Una catena di Markov irriducibile ammette sempre un'unica distribuzione di probabilità stazionaria.
- 2- Una volta che un sistema entra in uno stato transiente in una catena di Markov, rimane in tale stato indefinitamente

1- **Falso**. perché la catena raggiunge una distribuzione stazionaria solo se è ergodica, cioè irriducibile (tutti gli stati comunicano tra di loro), positiva e ricorrente (è possibile visitare gli stati precedentemente visitati con probabilità maggiore di 0). Una catena irriducibile non converge necessariamente ad una distribuzione stazionaria.

2- **Falso**. Uno stato S si definisce stato transiente se esiste uno stato J raggiungibile da S, ma S non è raggiungibile da J. Questa proprietà non implica che quando un sistema entra in uno stato transiente vi rimanga indefinitamente, anzi se un sistema che entra in uno stato vi rimane indefinitamente tale stato viene definito stato assorbente.

Domanda 1
Tempo rimasto 0:05:57

Risposta non ancora data

Punteggio max.: 4,00

▼
Contrassegna domanda

La **proprietà markoviana** è un concetto fondamentale nelle catene di Markov, che indica che lo stato presente di un sistema fornisce tutta l'informazione necessaria per prevedere lo stato futuro, indipendentemente dalla storia passata del sistema. In altre parole, la probabilità di transizione verso uno stato futuro dipende solo dallo stato corrente e non dai passi precedenti che hanno condotto a quel punto.

La **stazionarietà** implica che le probabilità di transizione e le probabilità di emissione rimangono costanti nel tempo, a condizione che lo stato corrente sia noto. Questa proprietà consente di semplificare ulteriormente i calcoli e l'analisi dell'HMM, poiché le probabilità di transizione e di emissione possono essere considerate costanti e non dipendono dal tempo o dalla posizione nella sequenza di osservazioni.

Esercitazione Catene di Markov

Esercizio 1

- Una industria produttrice di videoregistratori è talmente sicura della qualità dei propri prodotti da garantire per due anni la sostituzione dell'apparecchio in caso di guasto; le statistiche dell'azienda indicano che soltanto l'1% dei videoregistratori si guasta durante il primo anno ed il 5% durante il secondo. La garanzia non copre gli apparecchi forniti in sostituzione di quelli guasti. Formulare il problema come una catena di Markov, fornendo la matrice di transizione.

Dobbiamo identificare quindi le componenti della catena di Markov, iniziamo dagli **stati in cui può vertere il sistema che sono:**

- S0: videoregistratore nuovo
- S1: videoregistratore mai guasto nel primo anno
- S2: videoregistratore uscito da garanzia
- S3: videoregistratore guasto e sostituito

La matrice di transizione sarà 4x4:

$$P = \begin{bmatrix} 0 & 0.99 & 0 & 0.01 \\ 0 & 0 & 0.95 & 0.05 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Possiamo vedere come 0.01 è la probabilità del 1% data nella consegna. Conseguentemente anche il 0.99 nella prima riga identifica il videoregistratore che non si guasta nel primo anno. Allo stesso modo la seconda riga avrà la probabilità del 5% di essere guasto con 0.05 e il suo inverso 0.95. Le ultime due righe identificano la S2 e S3 che se un prodotto è in garanzia ci resta e che se un prodotto viene sostituito ci resta.

Esercizio 2

- Un processo produttivo è basato su di una macchina che si deteriora rapidamente così che alla fine di ogni giornata essa viene ispezionata per verificarne le condizioni. Queste ultime sono annotate e classificate in uno dei quattro seguenti stati:
 - stato 0: macchina come nuova
 - stato 1: macchina funzionante e poco deteriorata
 - stato 2: macchina funzionante ma molto deteriorata
 - stato 3: macchina danneggiata da sostituire con una nuova
- Il processo può essere modellato come una catena di Markov a stati finiti avente la seguente matrice di transizione:

$$P = \begin{bmatrix} 0 & \frac{7}{8} & \frac{1}{16} & \frac{1}{16} \\ 0 & \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Determinare lo steady state

$$\begin{cases} \pi_0 = p_{00}\pi_0 + p_{10}\pi_1 + p_{20}\pi_2 + p_{30}\pi_3 \\ \pi_1 = p_{01}\pi_0 + p_{11}\pi_1 + p_{21}\pi_2 + p_{31}\pi_3 \\ \pi_2 = p_{02}\pi_0 + p_{12}\pi_1 + p_{22}\pi_2 + p_{32}\pi_3 \\ \pi_3 = p_{03}\pi_0 + p_{13}\pi_1 + p_{23}\pi_2 + p_{33}\pi_3 \\ 1 = \pi_0 + \pi_1 + \pi_2 + \pi_3 \end{cases}$$

$\boxed{\pi_0 = \pi_2 = \pi_3 = 2/13}$
 $\pi_1 = 7/13;$

$$\begin{cases} \pi_0 = \pi_3 \\ \pi_1 = \frac{7}{8}\pi_0 + \frac{3}{4}\pi_1 \\ \pi_2 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_2 \\ \pi_3 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_2 \\ 1 = \pi_0 + \pi_1 + \pi_2 + \pi_3 \end{cases}$$

Avere uno steady state vuol dire che simulando dei passi in avanti si arriverà ad un punto in cui le probabilità della matrice non variano più. Quindi lo steady state è definito da K dove K è il numero di passi in avanti da fare prima che le probabilità della matrice non variano più. Per calcolare il valore di K bisogna impostare un sistema di N equazioni ad N incognite dove N è il numero di stati in cui può vertere il nostro sistema. Mettiamo quindi a π i vari stati in cui può vertere il nostro sistema e questi saranno uguali alla somma delle probabilità di partire da un qualunque stato del nostro sistema e arrivare allo stato che stiamo analizzando in questo momento. Quindi ad esempio per la prima riga mettiamo come primo elemento la probabilità di partire dallo stato 0 e raggiungere 0 moltiplicato per π dello stato da cui stiamo partendo (0)+ probabilità di partire dallo stato 1 e raggiungere 0 moltiplicato per π dello stato da cui stiamo partendo (1)+ e così via. Come ultima equazione si mette sempre $1 = \text{la somma di } \pi \text{ di ogni stato del sistema}$; in questo modo sto garantendo di avere una matrice stocastica. Per risolvere il sistema semplicemente sostituisco le probabilità di partire da uno stato e raggiungerne un altro con gli effettivi valori della matrice, in generale puoi andare a sostituire i valori degli elementi delle colonne che corrispondono allo stato che stai considerando al posto delle probabilità e lasciare i π senza toccarli. Vediamo come nell'esempio la prima riga vada tutta a 0 perché sto moltiplicando gli elementi della prima colonna per π che sono tutti 0 tranne l'ultimo che è 1. Una volta sostituito tutto risovi il sistema a N incognite e hai trovato tutti i valori dei π .

Esercizio 3

- Consideriamo la catena di Markov su $E = \{1, 2, 3, 4\}$ associata alla matrice di transizione

$$P = \begin{pmatrix} 0.25 & 0 & 0.75 & 0 \\ 0.25 & 0.5 & 0.25 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Esercizio 3

1. Qual è la probabilità partendo da 2 di essere in 2 dopo 2 passi?

$$p_{2,2}^{(2)} = (P^2)_{2,2} = \sum_h p_{2,h} \cdot p_{h,2} = \frac{1}{4}$$

$$P = \begin{pmatrix} 0.25 & 0 & 0.75 & 0 \\ 0.25 & 0.5 & 0.25 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

E partendo da 3?

$$p_{3,2}^{(2)} = (P^2)_{3,2} = \sum_h p_{3,h} \cdot p_{h,2} = 0$$

2. Qual è la probabilità di essere in 2 partendo da 2 dopo 12 passi?

$$p_{2,2}^{(12)} = 0.5^{12}$$

Qual è la probabilità partendo da 2 di essere in 2 dopo 2 passi?

Se non ci fosse stata la specifica di essere in 2 partendo da 2 allora avrei dovuto elevare al quadrato la matrice di transizione per fare una predizione a due step. In questo caso però basta elevare alla seconda soltanto la cella S2, S2 in quanto si sta partendo da S2 e si raggiunge S2, quindi moltiplicare gli elementi della seconda colonna per gli elementi della seconda riga. Se avessimo elevato alla seconda tutta la matrice e poi avessimo preso il valore in posizione S2, S2 avremmo ottenuto lo stesso risultato ma con molti più calcoli. Il risultato è quindi $0.25 \times 0 + 0.5 \times 0.5 + 0.25 \times 0 + 0 \times 1/3 = \frac{1}{4}$ perché sto moltiplicando i membri della seconda riga per i membri della seconda colonna

Se invece si partisse da S3 cambierebbe?

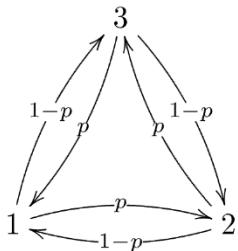
Si cambia, bisogna sempre elevare alla seconda ma la cella S3, S2 (il valore di partenza è la riga mentre quello di arrivo è la colonna) ottenendo quindi $0.5 \times 0 + 0 \times 0.5 + 0.5 \times 0 + 0 \times 1/3 = 0$. In pratica vado a calcolarmi la terza riga della tabella ottenuta elevando al quadrato la tabella originale e vedo in posizione 2 che valore c'è. Posso anche evitare di calcolare tutta la riga e calcolare semplicemente quello che c'è in posizione 2 facendo la moltiplicazione della riga 3 per la colonna 2.

Qual è la probabilità di essere in 2 partendo da 2 dopo 12 passi? Basta elevare alla 12 il valore S2,S2 = 0.5^{12}

Esercizio 4

- Consideriamo la catena di Markov sui vertici di un triangolo equilatero definita dalle regole seguenti: ad ogni istante essa si può spostare da un vertice a quello adiacente in senso antiorario con probabilità p e in senso orario con probabilità $1-p$, dove $0 < p < 1$.

a) Mostrare che la catena è regolare



$$P = \begin{pmatrix} 0 & p & 1-p \\ 1-p & 0 & p \\ p & 1-p & 0 \end{pmatrix}$$

Dopo aver rappresentato la catena e la sua matrice di transizione dobbiamo dire se la catena è regolare. La regolarità è determinata dall'irriducibilità, ovvero che ogni stato può raggiungere un qualsiasi altro stato della rete, e dalla stazionarietà, ovvero la presenza di un numero positivo tale che la matrice elevata a tal numero mi origina una probabilità >0 . Questo vuol dire che nella matrice finale non posso trovare al suo interno degli 0. In questo caso il valore è $K=2$ perché elevando alla seconda la matrice P otteniamo una nuova matrice con solo valori diversi da 0.

Esercizio 4

b) Calcolare per n grande le probabilità $P\{X_n = 1, X_{n+1} = 2\}$ e $P\{X_n = 2, X_{n+1} = 1\}$.

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j$$

$$P\{X_n = 1, X_{n+1} = 2\} = P\{X_{n+1} = 2 | X_n = 1\} \cdot P\{X_n = 1\} = p_{1,2} \cdot \pi_1 = \frac{p}{3}$$

$$P\{X_n = 2, X_{n+1} = 1\} = P\{X_{n+1} = 1 | X_n = 2\} \cdot P\{X_n = 2\} = p_{2,1} \cdot \pi_2 = \frac{1-p}{3}$$

Esercizio 5

1. Calcolare la probabilità di transizione a 2 step.

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix}$$

$$P^{(2)} = P \cdot P = \begin{bmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{bmatrix}$$

2. Se il sistema si trova nella modalità 1 alle 17:30, qual è la probabilità che esso si trovi nella modalità 1 alle 20:30?

$$P^{(3)} = P^{(2)}P = \begin{bmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{bmatrix} \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.496 & \dots \\ \dots & \dots \end{bmatrix}$$

Per ottenere il primo risultato si moltiplicano le righe della prima matrice per le colonne della seconda. In questo caso si fa per il primo elemento della matrice: $0.4 \cdot 0.4 + 0.6 \cdot 0.6 = 0.52$. Per il secondo elemento (il primo elemento della seconda riga) si fa $0.6 \cdot 0.4 + 0.4 \cdot 0.6 = 0.48$. Una volta trovati i primi due elementi della nuova matrice (ovvero la prima colonna della nuova matrice) si fa lo stesso per gli altri elementi, quindi si moltiplica gli elementi delle righe della prima matrice per gli elementi della seconda colonna della seconda matrice.

Per la seconda risposta invece faccio lo stesso ma posso evitare il calcolo intero della matrice dato che mi interessa solo lo stato 1 al tempo 3 quindi calcolo soltanto il primo elemento della prima colonna della matrice al tempo 3 ottenendo quindi $0.52 \cdot 0.4 + 0.48 \cdot 0.6 = 0.496$

Hidden Markov Model- ASSIGNMENT

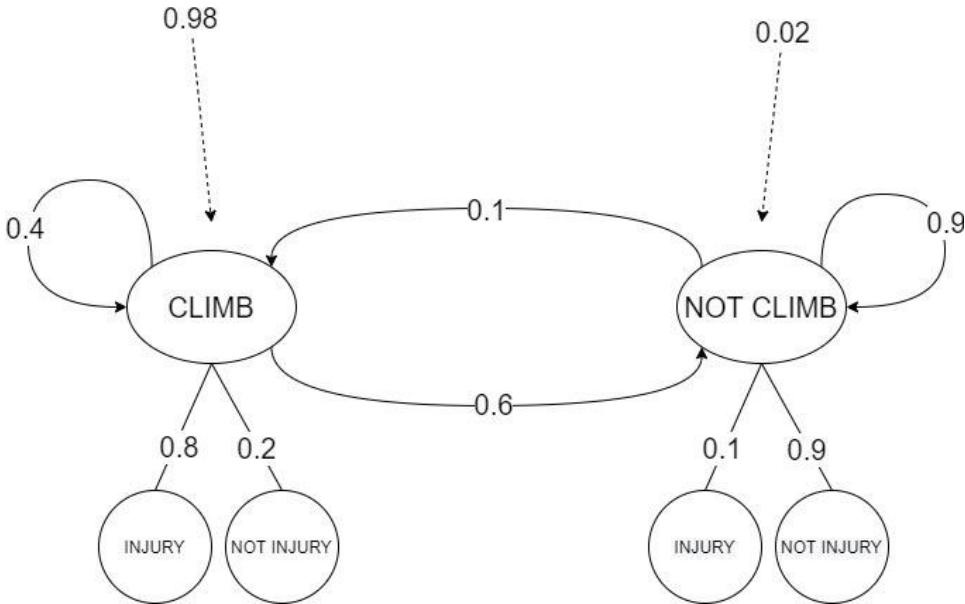
Si consideri il seguente HMM:

$$\pi = [0.98 \quad 0.02] \quad \text{Initial probability distribution (states: } climb, not climb\text{)}$$

$$T = \begin{bmatrix} 0.4 & 0.6 \\ 0.1 & 0.9 \end{bmatrix} \quad \text{Transition probability matrix (states: } climb, not climb\text{)}$$

$$O = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} \quad \text{Emission probability matrix (observation: } injury, not injury\text{)}$$

Determinare la struttura del modello grafico-probablistico.



Si effettui il task di filtraggio date le osservazioni: (injury, not injury).

Riportare il risultato nel box sottostante.

<0.11;0.89>

Per eseguire il task di filtraggio dobbiamo passare per le fasi di predizione e aggiornamento.

La fase di predizione serve per identificare il passaggio da uno stato temporale ad uno successivo e identificare la distribuzione di probabilità in quell'istante temporale. Per fare ciò devo fare la somma del prodotto di tutte le probabilità a priori degli stati, moltiplicate per la loro distribuzione delle probabilità in quel istante temporale.

In questo caso le due probabilità a priori sono 0.98 e 0.02 e le distribuzioni delle probabilità le otteniamo dalla matrice di transizione ottenendo $\langle 0.4; 0.6 \rangle$ e $\langle 0.1; 0.9 \rangle$. Facciamo quindi: $0.98 * \langle 0.4; 0.6 \rangle + 0.02 * \langle 0.1; 0.9 \rangle = \langle 0.392; 0.588 \rangle + \langle 0.002; 0.018 \rangle = \langle 0.394; 0.606 \rangle$

Ora passiamo alla fase di aggiornamento che serve per aggiornare la distribuzione delle probabilità sulla base dell'osservazione fatta all'istante T1, in questo caso injury. Per fare ciò in pratica moltiplico il risultato ottenuto dallo step di predizione per i valori della colonna della matrice delle osservazioni che identifica l'osservazione in analisi, in questo caso injury. Ricordati di mettere α in questa operazione e di calcolarlo come al solito per normalizzare i valori. Si avrà quindi $\langle 0.394; 0.606 \rangle * \langle 0.8; 0.1 \rangle \alpha = \langle 0.3152; 0.0606 \rangle \alpha$

Quindi $\alpha = 1 / 0.3152 + 0.0606 = 2.6609$, quindi $\langle 0.3152; 0.0606 \rangle \alpha = \langle 0.8387; 0.1613 \rangle = P(S1)$

Ora che abbiamo eseguito il filtraggio sulla prima osservazione facciamo lo stesso per la seconda. In questo caso la probabilità a priori che dobbiamo usare nella fase di predizione è il risultato ottenuto dal task di filtraggio eseguito in precedenza, ovvero $\langle 0.8387; 0.1613 \rangle$. Avremo quindi uno step di predizione che calcola: $0.8387 * \langle 0.4; 0.6 \rangle + 0.1613 * \langle 0.1; 0.9 \rangle = \langle 0.33548; 0.50322 \rangle + \langle 0.01613; 0.14517 \rangle = \langle 0.35161; 0.64839 \rangle$

Ora dobbiamo eseguire l'aggiornamento ma dobbiamo prestare attenzione all'osservazione estratta, in questo caso not injury, quindi dobbiamo prendere la colonna della matrice delle probabilità delle osservazioni che identifica not injury. Otteniamo quindi: $\langle 0.35161; 0.64839 \rangle * \langle 0.2; 0.9 \rangle \alpha = \langle 0.070322; 0.583551 \rangle \alpha$

Quindi $\alpha = 1 / 0.070322 + 0.583551 = 1.5293$, quindi $\langle 0.070322; 0.583551 \rangle \alpha = \langle 0.1075; 0.8925 \rangle = P(S2)$

Si calcoli la più probabile sequenza di stati usando l'algoritmo di Viterbi data la sequenza di osservazioni $E = \{\text{not injury}, \text{injury}, \text{not injury}\}$.

Scegli un'alternativa:

- a. <not climb, climb, not climb>
- b. <climb, climb, not climb> ✓
- c. <climb, not climb, not climb>
- d. < climb, climb, climb>
- e. <not climb, not climb, not climb>

Per prima cosa mi riprendo il grafico della catena di markov. Poi creo il grafico **di viterbi partendo dalla biforcazione degli stati Climb, Not Climb con le relative probabilità a priori 0.98 e 0.02**. Successivamente inizio a fare i riquadri delle varie osservazioni di E . Compilo i riquadri con la seguente formula: risultato dello step precedente * probabilità della matrice di transizione * probabilità dell'osservazione di E dato lo stato S . Devo prestare attenzione a quale probabilità di transizione usare in base alla freccia che sto considerando; se la freccia parte da un riquadro e va verso quello di un altro livello allora devo usare la probabilità di transizione dallo stato del riquadro di partenza allo stato del quadrato finale e allo stesso e nel caso in cui si resta nello stesso stato di prima allora devo vedere la probabilità di transizione di restare nello stesso stato. Devo prestare anche attenzione alla probabilità dell'osservazione dato lo stato in cui v'è che per ogni formula in un riquadro deve essere la stessa.

Esercitazioni Hidden Makrov Model

Esercizio 1

- Si supponga di volere modellare il comportamento di una semplice lampadina. Si consideri il tempo scandito in modo discreto (di secondo in secondo) e che i problemi che possono presentarsi siano i seguenti:
 - a) la lampadina può fulminarsi e smette di funzionare con probabilità 0.05.
 - b) la lampadina può risultare troppo calda, per cui ha difficoltà ad accendersi. La probabilità che la lampadina si possa surriscaldare troppo vale 0.15 e che in tale situazione abbia probabilità 0.35 di accendersi.

Si supponga inoltre che il sistema (perfetto) di rilevazione degli errori che si intende utilizzare per verificare il corretto funzionamento della lampadina indichi l'assenza di errori (good) o la presenza di errori (bad).

Modellare il problema mediante una catena di markov nascosta.

- **Insieme degli stati Ω .**

- “OK”: la lampadina non è fulminata
- “HOT”: la lampadina non è fulminata ma risulta surriscaldata.
- “KO”: la lampadina è fulminata.

$$\Omega = \{\text{OK}, \text{HOT}, \text{KO}\}$$

- **Insieme delle osservazioni Σ .**

- “GOOD”: non vengono rilevati errori
- “BAD”: vengono rilevati errori

$$\Sigma = \{\text{GOOD}, \text{BAD}\}$$

Un altro insieme che non vediamo è quello **della probabilità iniziale** che, assieme ai due insiemi precedenti, costituisce tutti gli insiemi che descrivono il problema:

- **Vettore di probabilità iniziale π .**

$$\pi = (1, 0, 0)^T$$

- **Matrice delle probabilità di emissione delle osservazioni E.**

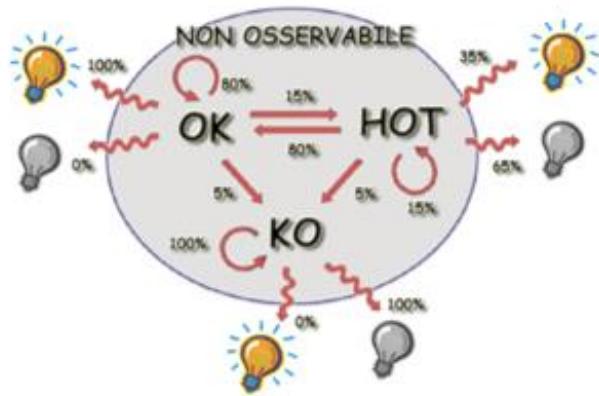
	GOOD	BAD
OK	1	0
HOT	0,35	0,65
KO	0	1

- **Matrice delle probabilità di transizione di stato A.**

	OK	HOT	KO
OK	0,80	0,15	0,05
HOT	0,80	0,15	0,05
KO	0	0	1

Il vettore della probabilità iniziale assumiamo che la lampadina quando la compriamo funzioni sempre. Per la matrice delle probabilità di transizione vediamo come il 0.05 è quella probabilità che fa passare la lampadina da OK e HOT a KO ovvero bruciata, 0.15 identifica la probabilità di surriscaldarsi, mentre il 0.80 è derivato dalla sottrazione delle due appena menzionate da 1. Per la matrice sensoriale abbiamo che da ognuno dei 3 stati in cui può vertere il sistema possiamo osservare o GOOD o BAD; essendo un sistema perfetto se siamo in uno stato OK allora può solo essere GOOD allo stesso modo ma inverso se è nello stato KO allora può solo essere BAD, infine per HOT sappiamo che 0.35 è la probabilità di accendersi quindi GOOD e per differenza troviamo anche BAD.

- Modello grafico probabilistico



In questa rappresentazione vediamo come la probabilità iniziale non viene evidenziata ma dovremmo fare delle linee tratteggiate che arrivano sullo stato OK, poi per la matrice di transizione basta fare il solito disegno, infine per la matrice delle

osservazioni abbiamo quelle linee con le onde che collegano gli stati alle lampadine.

Esercizio 2

- Dato un HMM caratterizzato da:
 - un insieme di stati $S=\{s_1, \dots, s_k\}$
 - un insieme di osservazioni $O=\{o_1, \dots, o_m\}$
- Quanti parametri è necessario specificare?



Esercizio 3

- Si consideri il seguente HMM caratterizzato da:

$$P = [0.5; 0.5] \text{ probabilità a priori}$$

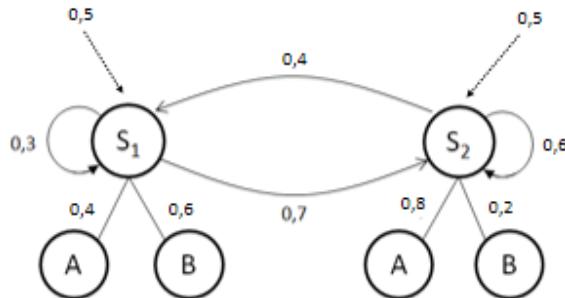
$$T = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix} \text{ probabilità di transizione}$$

$$O = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix} \text{ probabilità di emissione delle osservazioni}$$

Esercizio 3

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix} \quad O = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

- a) Determinare la struttura del modello grafico-probabilistico



Esercizio 3

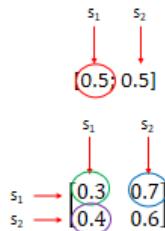
$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix} \quad O = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

- b) Calcolare la probabilità della sequenza di stati $S = \{s_1, s_1, s_2, s_1, s_2\}$

$$P(S|\theta) = \prod_{t=1}^{T-1} T_{t,t+1}$$

$$P(s_1, s_1, s_2, s_1, s_2) = \pi(s_1) \cdot P(s_1|s_1) \cdot P(s_2|s_1) \cdot P(s_1|s_2) \cdot P(s_2|s_1)$$

$$= 0.5 \cdot 0.3 \cdot 0.7 \cdot 0.4 \cdot 0.7 = 0.0294$$



Calcolare la probabilità di una sequenza di stati significa fare la produttoria delle probabilità dello stato che si vuole attraversare in un dato momento dato lo stato che lo precede. La probabilità del primo stato viene ottenuto dal vettore delle probabilità a priori, mentre i successivi vengono ottenuti dalla matrice delle transizioni avendo cura di utilizzare la riga che definisce lo stato di arrivo corretto.

Esercizio 3

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix} \quad O = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

- c) Calcolare la likelihood della sequenza di osservazioni $E = \{A, A, B, A, A\}$, data la sequenza di stati $S = \{s_1, s_1, s_2, s_1, s_2\}$

$$P(E|S, \theta) = \prod_{i=1}^{T-1} P(e_i|x_i, \theta)$$

$$P(A|s_1) \cdot P(A|s_1) \cdot P(B|s_2) \cdot P(A|s_1) \cdot P(A|s_2) =$$

$$0.4 \cdot 0.4 \cdot 0.2 \cdot 0.4 \cdot 0.8 = 0.01024$$



Per calcolare la likelihood bisogna calcolare la probabilità di avere la sequenza di osservazioni E data la sequenza di stati S e dalle osservazioni O quindi $P(E|S, O)$. Questo significa andare a calcolare la probabilità di emettere una certa osservazione sapendo che mi trovo in un determinato stato. Per fare questo usiamo la matrice delle osservazioni e vediamo la probabilità della combinazione di stato, identificato con le righe della matrice, e osservazione, identificato con le colonne della matrice.

Esercizio 3

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix} \quad O = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

d) Calcolare la joint likelihood della sequenza di osservazioni $E=\{A,A,B,A,A\}$ e la sequenza di stati $S=\{s_1,s_1,s_2,s_1,s_2\}$

$$\begin{aligned} P(E, S | \theta) &= P(E|S, \theta) \cdot P(S|\theta) \\ &= 0.01024 \cdot 0.0294 = 0.000301056 \end{aligned}$$

Per calcolare la **joint likelihood** bisogna calcolare la probabilità di $P(E, S | O) = P(E|S, O) * P(S|O)$. Il primo membro lo abbiamo già calcolato nello step precedente, mentre il secondo due esercizi fa. In questo modo calcoliamo la combinazione della probabilità della sequenza di stati S con le osservazioni E .

Esercizio 1

- Si consideri il seguente HMM che modella la probabilità di avere una annata molto calda/fredda, osservando la chioma degli alberi. Il modello HMM è caratterizzato dalle seguenti distribuzione di probabilità discrete:

- $P = [0.5; 0.5]$ probabilità a priori
- $T = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$ probabilità di transizione (stati: *hot*, *cold*)
- $O = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$ probabilità di emissione delle osservazioni (osservazioni: *small*, *large*)

Esercizio 1

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad E = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

a) Effettuare il task di filtraggio utilizzando le osservazioni *small*, *small*, *large*.

Al tempo $t=1$ osservo "small"

- Predizione da $t=0$ a $t=1$:

$$\begin{aligned} P(S_1) &= \sum_{s_1 \in S_0} P(S_1 | s_1) \cdot P(s_1) = \\ &= (< 0.7; 0.3 > * 0.5) + (< 0.4; 0.6 > * 0.5) = \\ &= < 0.35; 0.15 > + < 0.2; 0.3 > = < 0.55; 0.45 > \end{aligned}$$

perché stiamo avviando il sistema. Per fare ciò prendo la probabilità a priori dello stato S_1 ovvero 0.5 e la

Per fare il task di filtraggio dobbiamo utilizzare le osservazioni che ci vengono date e calcolarle una ad una iterando un processo di predizione e di aggiornamento.

Il passaggio di predizione consente di passare da un istante temporale ad un altro. Ovviamente per il primo passaggio avremo una partenza da t_0 a t_1

moltiplico per la distribuzione delle probabilità $<0.7;0.3>$ che identifica il possibile passaggio da S0 a S1 oppure il passaggio da S0 a S0, ovvero il passaggio da COLD a COLD e da COLD a HOT ovvero la prima riga della matrice di transizione. Ovviamente devo prendere anche in considerazione il caso in cui parto dallo stato COLD quindi eseguo lo stesso processo utilizzando la probabilità a priori di COLD ovvero 0.5 e la moltiplico per la distribuzione di probabilità identificata dalla seconda riga della matrice di transizione. Eseguo moltiplicazioni considerando le <> come parentesi e poi faccio la somma dei risultati finale. Il nuovo vettore identifica la probabilità di trovarsi in uno stato HOT o COLD nell'istante di tempo T1.

Esercizio 1

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad E = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

- a) Effettuare il task di filtraggio utilizzando le osservazioni *small, small, large*.

Al tempo t=1 osservo "small"

- Aggiorno con l'osservazione al tempo t=1 "small"

$$\begin{aligned} P(S_1|small) &= \alpha P(small|S_1) \cdot P(S_1) = \\ &= \alpha <0.4; 0.8> * <0.55; 0.45> = \\ &= \alpha <0.22; 0.36> = <0.379; 0.621> \end{aligned}$$

	small	large
HOT	0.4	0.6
COLD	0.8	0.2

Ricordati di mettere ALFA in questa operazione e di calcolarlo come al solito per normalizzare i valori.

Esercizio 1

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad E = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

- a) Effettuare il task di filtraggio utilizzando le osservazioni *small, small, large*.

Al tempo t=2 osservo "small"

- Predizione da t=1 a t=2:

$$\begin{aligned} P(S_2|S_1) &= \sum_{s_i \in S_1} P(S_2|s_i) \cdot P(s_i) = \\ &= (<0.7; 0.3> * 0.379) + (<0.4; 0.6> * 0.621) = \\ &= <0.265; 0.1137> + <0.2484; 0.3726> = <0.5134; 0.4863> \end{aligned}$$

variare dell'osservazione (small o large) risultato finale di ogni step identifica la probabilità di avere uno stato HOT o COLD all'istante temporale definito dallo step stesso.

L'operazione di aggiornamento aggiorna i valori con l'osservazione all'istante T1.

Quindi calcolo la probabilità di avere l'osservazione di analisi, in questo caso small, dato lo stato S1 moltiplicato per la probabilità di S1. Quindi in pratica moltiplico il risultato ottenuto dallo step di predizione per i valori della colonna della matrice delle osservazioni che identifica l'osservazione in analisi, in questo caso small.

I passaggi successivi al primo step di filtraggio completo sono analoghi solo che nello step di predizione si deve fare uso delle probabilità a priori ottenute dallo step di aggiornamento dell'istante temporale precedente. In sostanza quindi quello che varia da uno step ad un altro è solo la probabilità a priori e il valore delle probabilità delle osservazioni in fase di aggiornamento che varia al

Esercizio 1

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad E = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

a) Effettuare il task di filtraggio utilizzando le osservazioni *small, small, large*.

Al tempo $t=2$ osservo "small"

- o Aggiorno con l'osservazione al tempo $t=2$ "small"

$$\begin{aligned} P(S_2|small) &= \alpha P(small|S_2) \cdot P(S_2) = \\ &= \alpha <0.4; 0.8> * <0.5134; 0.4863> = \\ &= \alpha <0.20536; 0.38904> = <0.3455; 0.6545> \end{aligned}$$

Anche in questo caso, dato che l'osservazione è *small*, non cambia nulla rispetto allo step precedente.

Esercizio 1

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad E = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

a) Effettuare il task di filtraggio utilizzando le osservazioni *small, small, large*.

Al tempo $t=3$ osservo "large"

- o Predizione da $t=2$ a $t=3$:

$$\begin{aligned} P(S_3|S_2) &= \sum_{s_i \in S_2} P(S_3|s_i) \cdot P(s_i) = \\ &= (<0.7; 0.3>*0.3455) + (<0.4; 0.6>*0.6545) = \\ &= <0.241; 0.103> + <0.261; 0.392> = <0.505; 0.495> \end{aligned}$$

Qui abbiamo un'osservazione diversa dalle precedenti, ovvero *large*. In questo caso lo step di predizione resta il medesimo e vediamo che fa uso della distribuzione di probabilità calcolata dallo step di aggiornamento precedente. Per quanto riguarda invece lo step di aggiornamento vediamo come anziché fare uso della colonna della matrice delle osservazioni che identifica *small*, dobbiamo utilizzare i valori della colonna *large*.

Esercizio 1

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad E = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

a) Effettuare il task di filtraggio utilizzando le osservazioni *small, small, large*.

Al tempo $t=3$ osservo "large"

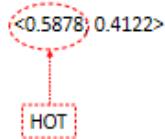
- o Aggiorno con l'osservazione al tempo $t=3$ "large"

$$\begin{aligned} P(S_3|large) &= \alpha P(large|S_3) \cdot P(S_3) = \\ &= \alpha <0.6; 0.2> * <0.505; 0.495> = \\ &= \alpha <0.303; 0.099> = <0.753; 0.247> \end{aligned}$$

Esercizio 1

b) Tenendo in considerazione il filtrato a 3 step, quale è la probabilità che il 5° anno sia "hot"?

$$P(S_5 | \text{small, small, large}) = \langle 0.753; 0.247 \rangle * T^2$$



Esercizio 2

- Dato il seguente HMM
 - $P = [0.5; 0.5]$ probabilità a priori
 - $T = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$ probabilità di transizione (stati: Q, R)
 - $O = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$ probabilità di emissione delle osservazioni (osservazioni: A, B)

Effettuare il task di smoothing considerando la seguente sequenza di osservazioni $E = \{A, A, B, A\}$

backward che va a calcolare la probabilità di emettere la sequenza dopo aver visitato un determinato stato nascosto.

Esercizio 2

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \quad O = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- Passo di Forward $\rightarrow P(x_k | e_{1:k})$
 - $f_0 = \langle 0.5; 0.5 \rangle$
 - $f_1 = \alpha f_0 T E_1 = \alpha \langle 0.5; 0.5 \rangle \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} = \langle 0.8182; 0.1818 \rangle$
 - $f_2 = \alpha f_1 T E_2 = \alpha \langle 0.8182; 0.1818 \rangle \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} = \langle 0.8834; 0.1166 \rangle$
 - $f_3 = \alpha f_2 T E_3 = \alpha \langle 0.8834; 0.1166 \rangle \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.1 & 0 \\ 0 & 0.8 \end{bmatrix} = \langle 0.1907; 0.8093 \rangle$
 - $f_4 = \alpha f_3 T E_4 = \alpha \langle 0.1907; 0.8093 \rangle \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} = \langle 0.7308; 0.2692 \rangle$

Il primo passaggio per il forward consiste nel copiare la probabilità a priori. Passiamo al filtraggio per il tempo 1 e andiamo a moltiplicare α per la distribuzione di probabilità dello step precedente, in questo caso la probabilità a priori, per la matrice di transizione per la matrice delle osservazioni dalla quale però andiamo a valorizzare solo la diagonale e i cui valori sono le colonne della matrice delle osservazioni originali

che fanno riferimento all'osservazione della sequenza che stiamo analizzando, in questo caso A. Per tutti gli step successivi il processo è lo stesso con l'unica differenza che la probabilità a priori è il risultato dello step precedente e che i valori della matrice delle osservazioni varia al variare di ciò che effettivamente viene osservato. Per eseguire il calcolo prima moltipichi le due matrici usando la stessa strategia di quando la elevi al quadrato, poi si moltiplica la matrice risultato per il vettore trasformando il vettore in una matrice ad una colonna e sulle righe i due valori del vettore. RICORDATI CHE IN OGNI PASSAGGIO C'E α DA CALCOLARE.

Esercizio 2

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \quad O = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- Passo di Backward $\rightarrow P(e_{k+1:t} | x_k)$

- $b_5 = <1; 1>$

- $b_4 = \alpha T E_4 b_5 = \alpha \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.6273 \\ 0.3727 \end{bmatrix}$

A

- $b_3 = \alpha T E_3 b_4 = \alpha \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.1 & 0 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} 0.6273 \\ 0.3727 \end{bmatrix} = \begin{bmatrix} 0.3695 \\ 0.6305 \end{bmatrix}$

B

- $b_2 = \alpha T E_2 b_3 = \alpha \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.3695 \\ 0.6305 \end{bmatrix} = \begin{bmatrix} 0.59 \\ 0.41 \end{bmatrix}$

A

- $b_1 = \alpha T E_1 b_2 = \alpha \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.59 \\ 0.41 \end{bmatrix} = \begin{bmatrix} 0.6465 \\ 0.3535 \end{bmatrix}$

A

Il secondo passaggio è quello di **backward**. Questo passaggio è caratterizzato da un ordine inverso rispetto a quello precedente, infatti prendiamo in considerazione le osservazioni dall'ultima fino alla prima. Il primo passo è sempre uguale e va messo come $<1;1>$ che identifica i passaggi che vengono dopo il tempo 4 di cui abbiamo le osservazioni. Passiamo al backward del tempo 4 e moltiplichiamo α per la matrice

di transizione per la matrice delle osservazioni, avendo cura di mettere sulla diagonale i valori dell'osservazione che stiamo analizzando e il resto a zero, per il vettore di backward ottenuto nello step precedente, in questo caso $<1;1>$ che è stato trasformato in una matrice ad una colonna con i due valori messi sulle righe. Otterremo come risultato una matrice ad una colonna con due righe valorizzate e questo risultato verrà usato nello step successivo come ultimo membro della moltiplicazione. Iteriamo il processo per ogni passaggio avendo cura di mettere i valori giusti nella matrice delle osservazioni e utilizzare il valore del risultato precedente come ultimo membro della moltiplicazione e abbiamo finito. RICORDATI CHE C'E SEMPRE α DA CALCOLARE.

Esercizio 2

$$P = [0.5; 0.5] \quad T = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \quad O = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- Passo di smoothing $\rightarrow P(x_k | e_{1:k}) * P(e_{k+1:t} | x_k)$

- $\delta_1 = \alpha f_1 b_2 = <0.8662; 0.1338>$

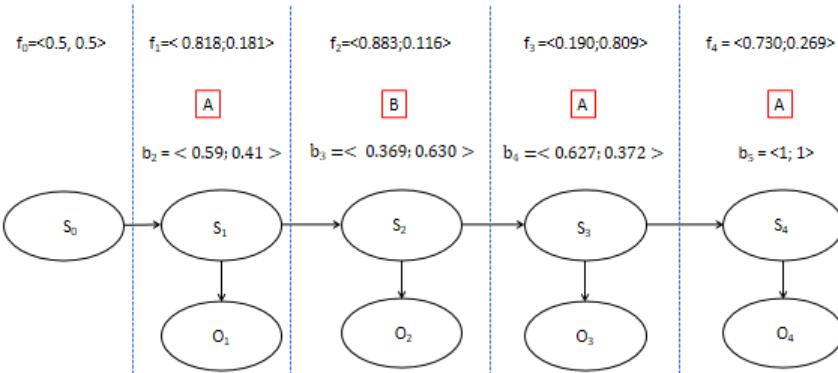
- $\delta_2 = \alpha f_2 b_3 = <0.8161; 0.1839>$

- $\delta_3 = \alpha f_3 b_4 = <0.2839; 0.7176>$

- $\delta_4 = \alpha f_4 b_5 = <0.7308; 0.2692>$

L'ultimo passaggio è quello di andare a moltiplicare il risultato del passaggio di forward di un istante temporale per il passaggio di backward dell'istante temporale successivo. Notare come l'ultimo step coincide con il risultato dell'ultimo step di forward dato che sto moltiplicando quel risultato per 1 che sarebbe il primo passaggio di backward.

Esercizio 2



Esercizio 3

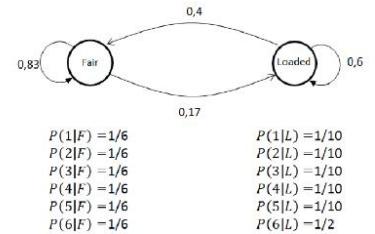
- Si consideri il seguente HMM che rappresenta l'alternanza tra un dado onesto (FAIR) ed un dado truccato (LOADED), le cui probabilità iniziali sono $P(\text{Loaded}) = 0.52$ e $P(\text{Fair} = 0.48)$.



$$\begin{aligned}
 P(1|F) &= 1/6 \\
 P(2|F) &= 1/6 \\
 P(3|F) &= 1/6 \\
 P(4|F) &= 1/6 \\
 P(5|F) &= 1/6 \\
 P(6|F) &= 1/6
 \end{aligned}
 \quad
 \begin{aligned}
 P(1|L) &= 1/10 \\
 P(2|L) &= 1/10 \\
 P(3|L) &= 1/10 \\
 P(4|L) &= 1/10 \\
 P(5|L) &= 1/10 \\
 P(6|L) &= 1/2
 \end{aligned}$$

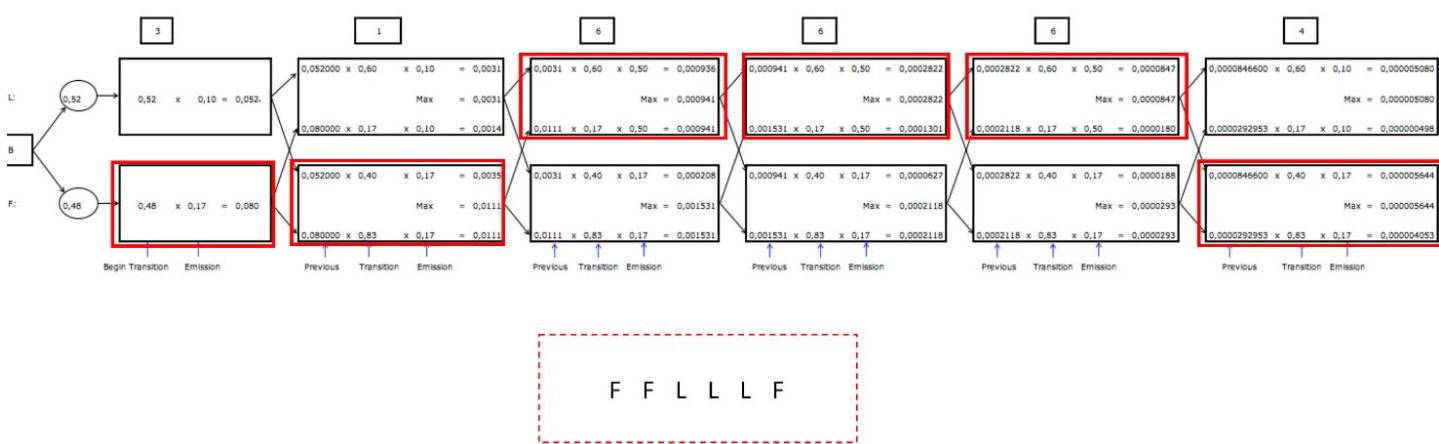
Data la sequenza di osservazioni $S = \{3, 1, 6, 6, 6, 4\}$, determinare la sequenza di stati più probabile utilizzando l'algoritmo di Viterbi.

In questo esercizio per il dado onesto tutte le facce hanno probabilità $1/6$ mentre per quello truccato c'è una probabilità di $1/2$ che esca 6.



$$\begin{aligned}
 P(1|F) &= 1/6 \\
 P(2|F) &= 1/6 \\
 P(3|F) &= 1/6 \\
 P(4|F) &= 1/6 \\
 P(5|F) &= 1/6 \\
 P(6|F) &= 1/6
 \end{aligned}
 \quad
 \begin{aligned}
 P(1|L) &= 1/10 \\
 P(2|L) &= 1/10 \\
 P(3|L) &= 1/10 \\
 P(4|L) &= 1/10 \\
 P(5|L) &= 1/10 \\
 P(6|L) &= 1/2
 \end{aligned}$$

Esercizio 3



Quando chiede di stimare la sequenza di stati più probabile data una sequenza di osservazioni bisogna usare sempre **Viterbi**. Per farlo iniziamo con andare a **simulare il processo di lancio dei dadi creando un grafico che vediamo sotto che simula tutti i percorsi possibili nei vari istanti temporali**. Per prima cosa abbiamo una biforcazione che identifica la probabilità iniziale ovvero 0.52 e 0.48 . Poi iniziamo ad **utilizzare le varie osservazioni per creare dei riquadri che identificano la probabilità di avere usato un dado normale o uno truccato**. Nella prima osservazione abbiamo **osservato 3** quindi nel quadrato del dado truccato dovrò moltiplicare la probabilità iniziale per la probabilità di osservare 3 sapendo che sto usando un dado truccato quindi $0.52 * 1/10$. Analogamente **per il dado non truccato** andrò a fare $0.48 * 1/6$. Per lo step successivo facciamo ancora due riquadri, vediamo che **osservo 1**, e **nei riquadri dovrò fare due calcoli**; uno che calcola la probabilità di rimanere nello stesso stato e di osservare 1 e uno che calcola la probabilità di passare da uno stato ad un altro e di osservare 1. Il calcolo quindi nel riquadro del dado truccato è di **moltiplicare il risultato dello step precedente per la probabilità di transizione dallo stato precedente a quello attuale per la probabilità di avere quella osservazione**. Ovviamente lo stato precedente può essere dado normale o dado truccato quindi vado ad inserire i numeri giusti: **$0.052 * 0.6 * 0.1 = risultato dello step precedente del dado$**

truccato*probabilità di passare dal dado truccato al dado truccato*probabilità di estrarre 1 dato che sto usando il dado truccato. Analogamente per il caso in cui passo da dado normale a dado truccato devo calcolare: $0.08*0.17*0.1 =$ risultato dello step precedente del dado normale* probabilità di passare dal dado normale al dado truccato* probabilità di estrarre 1 sapendo che sto usando un dado truccato.

Analogamente per il riquadro del dado normale faccio lo stesso ragionamento calcolando il caso in cui da dado normale resto in dado normale e estraggo 1 e il caso in cui da dado truccato passo a dado normale ed estraggo 1. La formula generale quindi è: **risultato stato precedente*probabilità della matrice di transizione*probabilità di osservazione dato lo stato di transizione**. Una volta fatti tutti i calcoli prendo da ogni riquadro il risultato più alto e lo tengo come risultato effettivo per calcolare i riquadri successivi. Faccio lo stesso processo per ogni osservazione che ho. Alla fine avrò per ogni riquadro una serie di probabilità e per vedere la sequenza più probabile mi basta scegliere PARTENDO DALL'ULTIMO RIQUEADRO E ANDANDO ALL'INDIETRO il riquadro con la probabilità maggiore e la freccia che definisce quella probabilità maggiore e quindi creare la sequenza di dadi utilizzati in base al riquadro selezionato.

TEORIA

Rete Bayesiana e D-Separazione

Illustrare come una rete bayesiana possa rappresentare una **distribuzione** congiunta in modo compatto e illustrare il concetto di **indipendenza condizionale anche in relazione alla D-separazione**.

Questo ragionamento può essere rappresentato tramite **una rete Bayesiana dove i nodi sono le variabili e gli archi stabiliscono una causalità diretta tra un nodo e l'altro**, in questo modo si riesce a **catturare le relazioni di indipendenza condizionata**:



La Rete Bayesiana è un grafo in cui i nodi sono annotati con una informazione quantitativa (tabelle di probabilità condizionata che quantificano la dipendenza tra variabili) che rappresenta le variabili casuali e gli archi orientati rappresentano una relazione di causalità diretta.

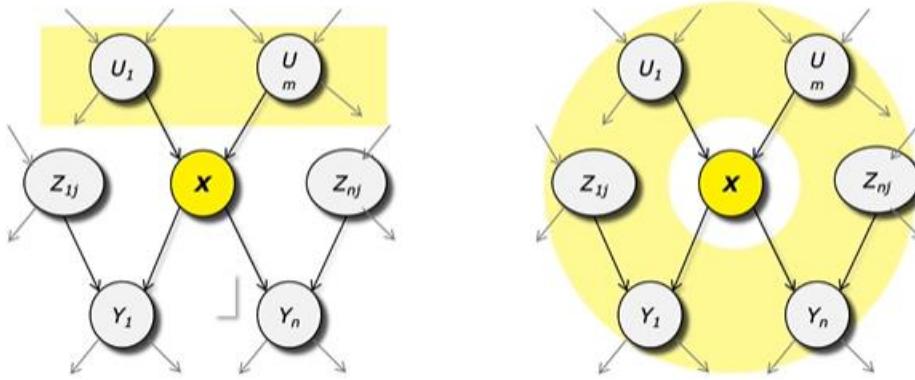
Nel contesto delle reti di Bayes, quest'idea di indipendenza condizionale è espressa attraverso l'uso dei nodi e degli archi della rete. Una rete di Bayes è costituita da nodi che rappresentano variabili casuali e da archi che rappresentano le relazioni di dipendenza tra queste variabili.

Quando si considerano due nodi collegati da un arco diretto, diciamo che i due nodi sono dipendenti condizionalmente. Ciò significa che il valore del nodo di partenza (genitore) influisce sulla distribuzione di probabilità del nodo di destinazione (figlio), dato il valore di altri nodi nella rete.

La **semantica topologica** viene fornita per mezzo di due specificazioni equivalenti che sono:

- 1- Un nodo è condizionalmente indipendente dai suoi non-descendenti (i nodi che stanno prima dei genitori) dati i suoi genitori.
- 2- Un nodo è condizionalmente indipendente da tutti i nodi restanti della rete, data la conoscenza dello stato dei suoi genitori, dei suoi figli e dei genitori dei suoi figli, ovvero l'insieme di nodi che è noto con il nome di **Markov Blanket**.

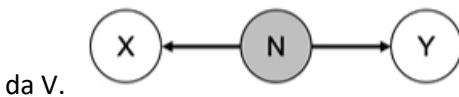
La seguente immagine spiega meglio i due punti:



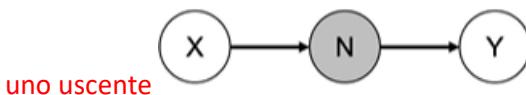
Il concetto di indipendenza condizionale di variabili casuali è collegato al concetto di D-separazione; questa è la proprietà che hanno due variabili di una rete quando queste sono condizionalmente indipendenti. Se due variabili sono condizionalmente indipendenti allora acquisire informazioni su una variabile non mi porta alcun beneficio di conoscenza sull'altra.

X e Y sono d-separate da un insieme E di variabili con evidenza (osservazioni) se e solo se **ogni cammino non orientato** da X a Y è “bloccato”. Dove un cammino è bloccato se e solo se vale almeno una delle tre seguenti condizioni:

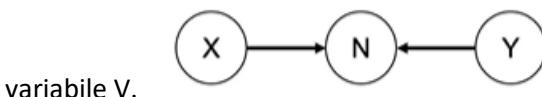
- 1- Lungo il cammino esiste una variabile V tale che appartiene all'insieme E delle variabili con evidenza e che gli archi che collegano V al cammino sono tail to tail, ovvero abbiamo archi uscenti



- 2- Lungo il cammino esiste una variabile V tale che appartiene all'insieme E delle variabili con evidenza e che gli archi che collegano V al cammino sono tail to head, ovvero un arco è entrante e uno uscente



- 3- Lungo il cammino esiste una variabile V tale che NON appartiene all'insieme E delle variabili con evidenza e che **NESSUNO** dei suoi discendenti appartiene all'insieme E delle variabili con evidenza e che gli archi che collegano V al cammino sono head to head, ovvero avere archi entranti nella variabile V.



Inferenza esatta

Inferenza per enumerazione - Metodo di eliminazione delle variabili

L'inferenza per enumerazione è un approccio per calcolare la distribuzione di probabilità di una variabile target specifica, data un'istanza di evidenze o osservazioni. Questo calcolo coinvolge la somma di prodotti di probabilità per tutte le possibili combinazioni di valori delle variabili coinvolte.

Il metodo di eliminazione delle variabili semplifica il calcolo dell'inferenza riducendo il numero di operazioni richieste. Il processo di eliminazione delle variabili coinvolge l'eliminazione graduale delle variabili non necessarie per calcolare la distribuzione di probabilità marginale del target.

I passaggi chiave del metodo di eliminazione delle variabili sono:

- 1- Identificazione dell'ordine delle variabili da eliminare

- 2- Propagazione delle probabilità attraverso la rete
- 3- Eliminazione delle variabili non necessarie per calcolare la distribuzione di probabilità marginale del target
- 4- Calcolo delle probabilità marginali usando i risultati delle eliminazioni precedenti

Le variabili che possono essere eliminate sono quelle che non influenzano direttamente la distribuzione della variabile di interesse una volta che le variabili di interesse sono state fissate.

Ad esempio, supponiamo di avere un modello con tre variabili: X, Y e Z. Vogliamo calcolare la distribuzione condizionale $p(X | Y = y)$. Nel processo di eliminazione delle variabili, possiamo eliminare la variabile Z se non ha un impatto diretto sulla distribuzione condizionale di X dato Y.

Inferenza nelle reti bayesiane - Regola di fattorizzazione

La regola di fattorizzazione, nota anche come regola di decomposizione o regola di decomposizione di Markov, è una regola fondamentale nelle reti bayesiane. Questa regola consente di esprimere la distribuzione di probabilità congiunta di tutte le variabili nel sistema come il prodotto delle distribuzioni di probabilità condizionate dei nodi in base alle loro dipendenze nella struttura della rete.

La regola di fattorizzazione nelle reti bayesiane afferma che la distribuzione di probabilità congiunta $P(X_1, X_2, \dots, X_n)$ di tutte le variabili X_1, X_2, \dots, X_n in una rete bayesiana può essere fattorizzata come:

$$P(X_1, X_2, \dots, X_n) = \prod P(X_i | \text{Genitori}(X_i))$$

Dove:

X_i rappresenta una variabile nel sistema.

$\text{Genitori}(X_i)$ rappresenta l'insieme dei genitori di X_i , cioè le variabili che hanno un arco diretto entrante nel nodo X_i nella struttura della rete.

4- Descrivere dettagliatamente l'algoritmo di inferenza esatta per enumerazione nelle reti Bayesiane

L'inferenza esatta per enumerazione, è un processo che serve per calcolare la distibuzione di probabilità di alcune variabili query, dato qualche evento osservato. In questa interrogazione, avremo le variabili X (variabili query), E (variabili prova) e Y (variabili nascoste).

Dato un evento I, le variabili di pro j e m, e le variabili nascoste t e a, il calcolo dell'inferenza esatta e la seguente:

$$P(I|j, m) = \alpha P(I, j, m) = \alpha \sum_t \sum_a P(I, t, a, j, m) \quad (2)$$

dove gli indici delle sommatorie, sono i possibili valori T e F delle variabili nascoste. Tale calcolo può essere semplificato, portando all'esterno della sommatoria i valori che rimangono costanti.

In sostanza quindi nel processo di inferenza, le variabili in evidenza avranno sempre un valore fisso, pertanto non è necessario considerare nel calcolo della distribuzione di probabilità i casi in cui queste assumano valori diversi. Le variabili non in evidenza però devono essere considerate per qualunque valore esse possano assumere, rendendo il metodo computazionalmente molto oneroso al crescere delle variabili non in evidenza.

Generazione numeri pseudo-casuali

I metodi computazionali per l'inferenza approssimata che vedremo poi forniscono dati tanto più affidabili quanto maggiore è il numero di eventi generati o campionati.

Una sequenza di numeri casuali è una sequenza di realizzazioni di variabili aleatorie indipendenti e identicamente distribuite. Quindi una **sequenza di numeri pseudo casuali** è una sequenza di numeri che sembrano **impredicibili**, ovvero da cui non si riesce ad estrarre alcuna regolarità.

Le proprietà statistiche delle serie di numeri pseudo casuali che vengono generati sono:

- 1- Indipendenza statistica
- 2- Uniformità della distribuzione
- 3- Riproducibilità della sequenza di valori
- 4- Non ripetitività su un prefissato periodo

Generazione numeri pseudo-casuali - Metodo di trasformazione inversa

Andando nel dettaglio della **tecnica di trasformazione inversa** possiamo dire che se si vuole generare una variabile aleatoria X con funzione di densità di probabilità $f_X(x)$ allora si deve:

- 1- Calcolare la **funzione di distribuzione di probabilità** o funzione cumulativa di probabilità con la

$$F_X(x) = \int_{-\infty}^x f_X(\tau) d\tau$$

seguinte formula: tale funzione è continua, monotona crescente ed è sempre compresa tra 0 e 1.

- 2- Si pone $u=F_X(x)$ con u un numero random compreso tra 0 e 1

3. si risolve:

$$X = F_X^{-1}(u)$$

ottenendo che la variabile aleatoria X è distribuita secondo $f_X(x)$ ($X \sim f_X(x)$)

Ripetendo questi step tante volte si **otterrà la voluta distribuzione di numeri casuali**.

Generazione numeri pseudo-casuali - Acceptance rejection/ accettazione rifiuto

Vediamo ora il **metodo di accettazione/rifiuto**. Questo metodo consente di generare una sequenza di numeri pseudo casuali **senza utilizzare il calcolo della funzione inversa F^{-1}** tipico della tecnica di trasformazione inversa.

Questo metodo viene applicato nel **caso in cui abbiamo delle distribuzioni definite su intervalli finiti $[a,b]$** .

Conoscendo la densità di probabilità $f_x(x)$ definita su un intervallo finito $[a, b]$ e la sua immagine nel codominio $[0, c]$, si generano delle sequenze pseudo-casuali u_1 e u_2 su $[0,1]$;

si derivano altre due sequenze uniformi con $\begin{cases} x = a + (b - a) u_1 \\ y = c u_2 \end{cases}$;
se la coppia (x, y) ricade all'interno di $F_x(x)$ viene accettata.

È efficiente quando l'area di $F_x(x)$ copre quasi tutto $[a, b] \times [0, c]$, altrimenti si scartano troppe coppie.

Inferenza approssimata

Inferenza approssimata - Direct sampling/campionamento diretto

Ogni variabile viene campionata in ordine topologico con una distribuzione dipendente dal condizionamento risultante sui nodi genitore, in modo efficiente. La stima è consistente, cioè converge verso il valore reale quando il numero di campioni tende a ∞ .

Inferenza approssimata - Rejection sampling/ campionamento con rigetto

Se volessi computare la probabilità condizionata dove esistono delle variabili con evidenza, allora devo utilizzare un algoritmo diverso, ovvero il **campionamento con rigetto**. Questo genera campioni come il campionamento diretto e, successivamente, rifiuta tutti quei campioni generati che non sono conformi dal punto di vista dell'evidenza. In sostanza quindi la probabilità stimata di un certo valore X della distribuzione date le evidenze sarà uguale a: $P(X|E=e)$. Il problema principale di questo algoritmo sono che al crescere delle variabili con evidenze, cresceranno di conseguenza i campioni rigettati, quindi se in una rete ci sono multiple evidenze allora l'algoritmo NON può essere utilizzato.

Inferenza approssimata - Descrivere l'algoritmo likelihood weighting

Un'alternativa all'algoritmo di campionamento con rigetto che presenta problemi nel caso di multiple evidenze è l'algoritmo **di Likelihood Weighting**, che, invece di generare dei campioni che verranno poi rigettati, pesa i campioni rispetto alla likelihood dell'evento. L'algoritmo fissa il valore delle variabili per le quali è disponibile evidenza ($E=e$) e, successivamente, genera campioni solo per i nodi restanti X e Y . Il peso che l'algoritmo attribuisce ai vari eventi (i campioni generati) è la likelihood che l'evento associa all'evidenza.

L'algoritmo **utilizza tutti i campioni** generati a differenza del campionamento per rigetto, però al crescere delle evidenze le sue performance diminuiscono in quanto **molti dei campioni estratti avranno un peso molto piccolo**.

Algoritmo che fissa il valore delle variabili con evidenza, poi genera campioni solo per le restanti. Pesa i diversi eventi con un valore che corrisponde alla verosimiglianza associata all'evidenza. Alla fine la probabilità query viene stimata con un rapporto tra le verosimiglianze del valore ricercato e le verosimiglianze totali.

$$w_{s1} = P(e_1|x)P(e_2|x) \dots \\ w_{s2} = P(e_1|\bar{x})P(e_2|\bar{x}) \dots \\ P(q|e_1, e_2) = \frac{\sum_{(s_i:q)} w_{s_i}}{\sum_{s_i} w_{s_i}}$$

Inferenza approssimata - Markov Chain Monte Carlo MCMC

1-Desrivere Markov Chain Monte Carlo per il task di inferenza (approssimata) nelle Reti Bayesiane.

A differenza degli altri due algoritmi (likelihood weithing e rejection sampling), l'algoritmo MCMC (Markov Chain Monte Carlo), non è basato sul campionamento creando da zero gli eventi, ma ad ogni evento apporta una modifica casuale a quello precedente. La rete è associato uno stato corrente, e lo stato successivo è generato camminando casualmente lungo la rete, e modificando la variabile non di prova X_i .

MCMC si può dire che "cammina" casualmente nello stato degli stati modificando una variabile per volta, ma tenendo fisso il valore delle variabili di prova.

Invece di partire da zero per generare gli eventi, modifica casualmente quelli già elaborati. Ragiona come se il modello di rete Bayesiana avesse uno stato corrente, dato dall'assegnamento di valori a ciascun nodo. Lo stato successivo si ottiene campionando una alla volta le variabili senza evidenza.

$$P(x_i|mb(x_i)) = \propto P(x_i|genitori(x_i)) \cdot \prod_{y \in figli(x_i)} P(y_i|genitori(y_i))$$

Lo stato iniziale si ottiene fissando le evidenze e inizializzando casualmente le variabili senza evidenza; col campionamento si ottiene una nuova distribuzione poi usata per generare lo stato successivo.

Catena di Markov - Definizione di catena di Markov

Una catena di Markov è un processo stocastico che **descrive un sistema in cui un'entità o uno stato cambia nel tempo in accordo con una proprietà chiamata proprietà di Markov**. Una catena di Markov è caratterizzata dall'assenza di memoria: lo stato futuro dipende solo dallo stato corrente e non è influenzato dagli stati precedenti.

Formalmente, una catena di Markov è definita da un insieme di stati $S = \{S_1, S_2, \dots, S_n\}$ e una matrice di transizione T , in cui ogni elemento $T(i, j)$ rappresenta la probabilità di transizione dallo stato S_i allo stato S_j .

Una catena di Markov per essere definita correttamente deve avere: un insieme di stati S , la probabilità di transizione tra stati e la distribuzione iniziale degli stati.

Proprietà dei processi Markoviani - Steady state impostare il sistema

Avere uno steady state vuol dire che simulando dei passi in avanti si arriverà ad un punto in cui le probabilità della matrice non variano più. **Quindi lo steady state è definito da K dove K è il numero di passi in avanti da fare prima che le probabilità della matrice non variano più.**

Vettore contenente le probabilità a lungo termine di trovarsi in ciascuno stato: $\lim_{t \rightarrow \infty} P_{ij}(t) = \pi_j$

Si calcola risolvendo il sistema $(T - I|e)^T P_i = b$, oppure: traspongo la matrice di transizione, pongo ciascuna riga $= \pi_j$ e l'ultima come somma $\pi_1 + \dots + \pi_n = 1$. Posso rimuovere una riga per risolvere più facilmente.

Proprietà dei processi Markoviani

Catena **STAZIONARIA**: Una catena di Markov è stazionaria se la sua distribuzione di probabilità non cambia nel tempo. In altre parole, una volta raggiunto l'equilibrio, la distribuzione di probabilità degli stati rimane costante indipendentemente dal tempo trascorso.

Catena **IRRIDUCIBILE**: Una catena di Markov è irriducibile se è possibile raggiungere qualsiasi stato partendo da qualsiasi altro stato in un numero finito di passi. In altre parole, ogni stato della catena è accessibile da ogni altro stato.

Catena **ERGODICA**: Una catena di Markov ergodica è simultaneamente irriducibile, ricorrente e positiva. Questo significa che la catena può raggiungere qualsiasi stato partendo da qualsiasi altro stato, ritorna a visitare gli stati con una certa frequenza nel lungo termine. Posiamo quindi dire che per le catene ergodiche è garantito che verrà raggiunta una distribuzione di equilibrio, ovvero raggiungerà uno steady state ad un certo punto che è uno stato che non varia più.

Catena **REGOLARE**: La regolarità è determinata dall'irriducibilità, ovvero che ogni stato può raggiungere un qualsiasi altro stato della rete, e dalla aperiodicità, ovvero non presenta cicli di lunghezza fissa.

Catena a comportamento **TRANSITORIO**: Uno stato transitorio in una catena di Markov è uno stato in cui c'è una probabilità positiva di lasciare lo stato e passare ad altri stati, quindi quando la catena non ha ancora raggiunto l'equilibrio.

Stato **RAGGIUNGIBILE**: Uno stato è raggiungibile in una catena di Markov se esiste una sequenza di transizioni che può portare dalla partenza allo stato di arrivo.

Stato **ASSORBENTE**: uno stato assorbente è uno stato da cui non si può uscire una volta raggiunto. In altre parole, una volta che un processo raggiunge uno stato assorbente, rimarrà in quel determinato stato per il resto del tempo. Ha $P_{ii} = 1$, Gli stati assorbenti si riconoscono facilmente perché ad ognuno di essi corrisponde, nella matrice di transizione, un elemento sulla diagonale principale eguale ad 1 (questo e, necessariamente l'unico elemento differente da zero della sua riga).

Stato **TRANSIENTE***: se $\exists j$ raggiungibile da i , ma i non è raggiungibile da j . Proprietà di classe: se uno stato ce l'ha, tutti gli stati con cui comunica ce l'hanno.

Stato **RICORRENTE***: se non è assorbente né transiente. Uno stato ricorrente in una catena di Markov è uno stato che verrà visitato infinitamente volte nel corso del tempo. Proprietà di classe: se uno stato ce l'ha, tutti gli stati con cui comunica ce l'hanno.

Stato **PERIODICO** di periodo K : se $\exists K > 1$: tutti i cammini che dallo stato i tornano in i hanno lunghezza pari a un multiplo di K . Ciò significa che la catena deve tornare a uno stato periodico dopo un certo numero di transizioni.

Hidden Markov Models - Definisci cosa si intende per proprietà Markoviana

La proprietà markoviana è un concetto fondamentale nelle catene di Markov, che indica che lo stato presente di un sistema fornisce tutta l'informazione necessaria per prevedere lo stato futuro, indipendentemente dalla storia passata del sistema. In altre parole, la probabilità di transizione verso uno stato futuro dipende solo dallo stato corrente e non dai passi precedenti che hanno condotto a quel punto.

Hidden Markov Models - Definisci una catena di Markov a stati nascosti HMM

Processo stocastico continuo o discreto caratterizzato da insieme degli stati X_t , insieme delle osservazioni E_t , matrice delle probabilità di transizione $P(X|X_{t-1})$, matrice di probabilità di emissione delle osservazioni $P(E_t|X_t)$, matrice delle probabilità iniziali degli stati $P(x_0)$.

MODELLO DI TRANSIZIONE $P(X|X_{0:t-1}) = P(X|X_{t-1})$ descrive la fisica di moto

MODELLO SENSORIALE $P(E_t|X_{0:t-1}, E_{0:t-1}) = P(E_t|X_t)$ descrive il processo di misurazione

Una catena di Markov a stati nascosti (**Hidden Markov Model**, HMM) è un modello statistico che combina una catena di Markov con un processo generativo di osservazioni. In un HMM, i "veri" stati del sistema, chiamati stati nascosti, non sono direttamente osservabili, ma possono essere dedotti o stimati a partire dalle osservazioni disponibili.

Un HMM è composto da:

- 1- Insieme degli stati
- 2- Insieme delle osservazioni
- 3- Probabilità a priori
- 4- Matrice di transizione degli stati
- 5- Matrice sensoriale delle osservazioni

Hidden Markov Models – Inferenza

Hidden Markov Models inferenza - Descrivi il task di filtraggio con HMM

Il task di filtraggio consiste nel calcolare la distribuzione di probabilità degli stati nascosti correnti dato l'insieme delle osservazioni fino a quel momento. L'algoritmo di filtraggio è composto da due step principali: la proiezione e l'aggiornamento. Supponiamo di avere il filtraggio al tempo t e di volere eseguire il filtraggio al tempo $t+1$. Per fare ciò innanzitutto eseguo una proiezione portando avanti la distribuzione dello stato corrente da t a $t+1$ utilizzando il modello di evoluzione che in questo caso è una catena di Markov. Dopo aver eseguito la proiezione devo aggiornare la previsione rispetto all'evidenza dello stato $t+1$. Vediamo ora nel dettaglio le formule che portano all'esecuzione del filtraggio:

Step di predizione:

$$\begin{aligned}
 P(X_{t+1} | e_{1:t+1}) &= P(X_{t+1} | e_{1:t}, e_{t+1}) && \text{divido l'evidenza} \\
 &= \alpha P(e_{t+1} | X_{t+1}, e_{1:t}) P(X_{t+1} | e_{1:t}) && \text{uso il teorema di Bayes} \\
 &= \alpha P(e_{t+1} | X_{t+1}) P(X_{t+1} | e_{1:t}) && \text{uso la proprietà di Markov} \\
 &\quad \text{dell'evidenza}
 \end{aligned}$$

- o α è la costante di normalizzazione della somma di probabilità

Step di aggiornamento:

$$\begin{aligned}
 P(X_{t+1} | e_{1:t+1}) &= \alpha P(e_{t+1} | X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t, e_{1:t}) P(x_t | e_{1:t}) \\
 &= \alpha P(e_{t+1} | X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t) P(x_t | e_{1:t}) && \text{uso la proprietà di Markov} \\
 f_{1:t+1} &= \alpha \text{Forward}(f_{1:t}; e_{t+1})
 \end{aligned}$$

3-Scrivere il task di Filtraggio e di Previsione

Il filtraggio calcola la distribuzione a posteriori dello stato corrente, date tutte le prove raccolte. In altre parole si vuole calcolare la $P(X_t | e_{1:t})$. Dal filtraggio si può facilmente calcolare il risultato di $t + 1$ dalla nuova $e + 1$. Per far ciò, si proietta in avanti da t a $t+1$, e la si aggiorna in base al nuovo $e + 1$.

$$\begin{aligned}
 P(X_{t+1} | e_{1:t+1}) &= P(X_{t+1} | e_{1:t}, e_{t+1}) && \text{(dividendo le prove)} \\
 &= \alpha P(e_{t+1} | X_{t+1}, e_{1:t}) P(X_{t+1} | e_{1:t}) && \text{(per la regola di Bayes)} \\
 &= \alpha P(e_{t+1} | X_{t+1}) P(X_{t+1} | e_{1:t}) && \text{(per la proprietà di Markov della prova)}
 \end{aligned}$$

Per ottenere la predizione, dello stato corrente X_t :

$$\begin{aligned}
 P(X_{t+1} | e_{1:t+1}) &= \alpha P(e_{t+1} | X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t, e_{1:t}) P(x_t | e_{1:t}) \\
 &= \alpha P(e_{t+1} | X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t) P(x_t | e_{1:t}) && \text{(sfrutta la proprietà di Markov)}
 \end{aligned}$$

Nella sommatoria il primo termine non è altro che il modello di transizione, il secondo la distribuzione dello stato corrente. Implementiamo il tutto sotto una forma ricorsiva, ottenendo la seguente formula:

$$f_{1:t+1} = \alpha \text{Forward}(f_{1:t}, e_{t+1}) \tag{1}$$

Descrivi i task di inferenza con HMM

5-Elencare e descrivere brevemente i task di inferenza in un HMM

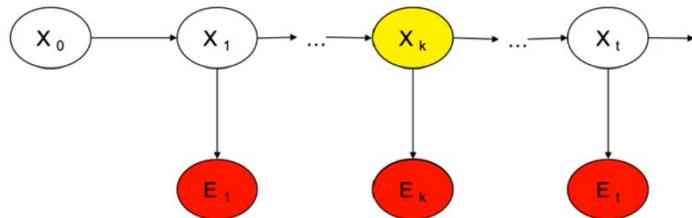
I task per l'inferenza su modelli temporali HMM, sono i seguenti:

- Filtraggio: calcola la distribuzione dello stato corrente, avendo le osservazione fino a quel momento
- Predizione: calcola la distribuzione a posteriori di un elemento futuro avendo a disposizione i dati fino a un determinato momento. Calcola $P(X_{t+k}|e_{1:t})$ per un $k > 0$.
- Smoothing: effettua una regolazione di uno stato passato a posteriori. Calcola $P(X_k|e_{1:t})$ per un qualsiasi k tale che $0 \leq k < t$.
- Spiegazione più probabile: data una sequanza di osservazione, calcola la sequenza di stati che più probabilmente ha generato tali osservazioni. Si vuole calcolare $\text{argmax}_{x_{1:t}} P(x_{1:t}|e_{1:t})$

Hidden Markov Models inferenza - Descrivere Task di smoothing

Lo smoothing è il task che permette di rivedere la distribuzione di probabilità di uno stato al tempo $t-k$ avendo acquistato le informazioni fino al tempo t . Più formalmente è il processo di calcolo della distribuzione di stati passati, date le osservazioni fino allo stato corrente.

$$P(X_k|e_{1:t}) \text{ per } 1 \leq k < t$$



Per effettuare lo smoothing **dobbiamo considerare separatamente le osservazioni che precedono k e quelle che succedono k** . Bisogna quindi applicare una procedura **filtraggio forward da 1 fino a k** e poi una procedura **backward da t fino a k** .

Calcolo della distribuzione a posteriori di stati passati date le osservazioni finora.

$$\propto P(X_k|e_{1:k})P(e_{k+1:t}|X_k)$$

Passo forward/filtraggio: $f_i = \propto f_{i-1} T E_i$

Passo backward: $b_j = \propto T E_j b_{j+1}$

f_0 è la distribuzione a priori, poi aggiorno in ordine crescente;

T è la matrice di transizione fra stati;

E_n è la forma matriciale dell'osservazione n -esima, ottenuta mettendo sulla diagonale la colonna di 0 corrispondente all'osservazione corrente, e il resto a zero;

b_k è un vettore fittizio di soli 1, poi aggiorno in ordine decrescente.

2-Scrivere dettagliatamente il task di Smoothing

Lo Smoothing o regolarizzazione è il calcolo della distribuzione a posteriori di uno stato passato, date tutte le prove raccolte. In pratica si vuole calcolare la probabilità $P(X_k|e_{1:t})$ per qualche k tale che $0 \leq k < t$. Nell'esempio dell'ombrellino, è calcolare la probabilità che abbia piovuto mercoledì scorso, date tutte le osservazioni fatte fino a quel momento. La regolarizzazione fornisce una stima dello stato migliore, per via del maggior numero di prove raccolte.

Il processo migliore è quello di separare le prove, fino ad arrivare a k , e da $k+1$ fino a t .

$$\begin{aligned} P(X_k|e_{1:t}) &= P(X_k|e_{1:k}, e_{k+1:t}) \\ &= \alpha P(X_k|e_{1:k})P(e_{k+1:t}|X_k, e_{1:k}) && \text{(usando la regola di Bayes)} \\ &= \alpha P(X_k|e_{1:k})P(e_{k+1:t}|X_k) && \text{(usando l'indipendenza condizionale)} \\ &= \alpha f_{1:k} b_{k+1:t} \end{aligned}$$

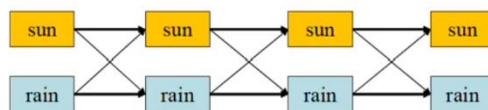
$f_{1:k}$ può essere calcolato filtrando in avanti da 1 a k . $b_{k+1:t}$, può essere calcolato con un processo ricorsivo, che va all'indietro da t :

$$\begin{aligned} P(e_{k+1:t}|X_k) &= \sum_{x_{k+1}} P(e_{k+1:t}|X_k, x_{k+1})P(x_{k+1}|X_k) && \text{(condizionando per } X_{k+1}) \\ &= \sum_{x_{k+1}} P(e_{k+1:t}|x_{k+1})P(x_{k+1}|X_k) && \text{(per l'indipendenza condizionale)} \\ &= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) \\ &= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k) \end{aligned}$$

Hidden Markov Models inferenza - Relazione ricorsiva su cui si basa algoritmo di Viterbi

Per spiegare la sequenza più probabile possiamo dire che questa è, **data una sequenza di osservazioni, la sequenza di stati che più probabilmente ha generato il mio set di osservazioni**.

Questi algoritmi vanno ad applicare delle soluzioni a dei grafi e quindi è necessario **trasformare il problema della sequenza più probabile in un problema su un grafo**. Per fare ciò si prendono tutti i valori che uno stato può assumere in un determinato istante e quelli diventano i nodi del grafo. **Gli archi invece identificano la probabilità di transizione da uno stato ad un altro**.



La probabilità di ogni cammino poi viene calcolata come il prodotto delle probabilità di transizione (ottenute dal modello di transizione) **per le probabilità delle osservazioni rilevate in ogni stato** (ottenute dal modello sensoriale). Per risolvere il problema si utilizza **l'algoritmo di Viterbi**:

Si basa sull'assunzione che esista una relazione ricorsiva tra i cammini più probabili verso ogni stato x_{t+1} e i cammini più probabili verso ogni stato x_t , che permette di ottenere la probabilità della sequenza più probabile per raggiungere ogni stato finale.

$$\max_{x_1, \dots, x_t} P(x_1, \dots, x_t, x_{t+1}|e_{1:t+1}) = \alpha P(e_{t+1}) \max_{x_t} (P(x_{t+1}|x_t) \max_{x_1, \dots, x_{t-1}} (P(x_1, \dots, x_{t-1}|e_{1:t}))$$

Costruisco un grafico Trellis prendendo sempre la probabilità massima precedente \times la probabilità di transizione \times la probabilità di osservazione, poi ripercorro all'indietro per la sequenza più probabile.

Filtri di Kalman

Illustrare i principi del filtro di kalman

Il filtro di Kalman è un algoritmo di stima ottimale che permette di combinare le informazioni misurate con un modello matematico del sistema per ottenere stime accurate dello stato del sistema in presenza di rumore e incertezza. Si basa su principi di previsione e correzione, utilizzando le osservazioni disponibili per migliorare l'accuratezza delle stime dello stato del sistema. L'obiettivo finale del filtro di Kalman è ottenere stime accurate e ottimali dello stato del sistema, riducendo l'effetto del rumore e dell'incertezza presenti nelle osservazioni.

Nelle catene nel filtro di Kalman sapete che l'assunzione sulle gaussiane dei modelli sensoriali, dei modelli di transizione

Le assunzioni gaussiane nei filtri di Kalman si riferiscono all'ipotesi che le distribuzioni di probabilità delle variabili coinvolte nel filtro di Kalman siano distribuzioni gaussiane. Questa assunzione di gaussianità semplifica notevolmente il calcolo e l'implementazione del filtro di Kalman, permettendo di utilizzare metodi analitici e riducendo la complessità computazionale.

Le assunzioni gaussiane nei filtri di Kalman includono:

- 1- Assunzione di normalità
- 2- Assunzione di linearità
- 3- Assunzione di omoschedasticità
- 4- Assunzione di indipendenza: