0.1 Probabilità

La probabilità è una misura che permette di quantificare la possibilità che un evento si verifichi oppure no. Ha lo scopo di formulare e studiare modelli matematici che permettono di descrivere il comportamento di un evento, nonostante non si sia in grado di predirlo con totale certezza. Secondo la definizione classica la probabilità di un evento casuale e' data dal rapporto tra il numero di casi favorevoli al presentarsi di tale evento ed il numero totale di casi possibili, supponendo che tutti i possibili casi siano equiprobabili. In termini matematici:

$$P(E) = \frac{n_E}{n} \tag{1}$$

dove n_E corrisponde al numero di volte in cui potrebbe verificarsi E ed n è il numero di esperimenti fatti. Tuttavia è possibile riconoscere due diverse interpretazioni della probabilità:

- interpretazione frequentista: la probabilità viene definita in modo oggettivo come il rapporto tra il numero di volte in cui si verifica l'evento di interesse ed il numero totale di prove effettuate, per un numero di prove che tende all'infinito. In termini matematici:

$$P(E) = \lim_{x \to \infty} \frac{n_e}{n} \tag{2}$$

- interpretazione bayesiana: la probabilità viene definita in modo soggettivo, infatti alle ipotesi viene associata una probabilità che indica il grado di fiducia che un individuo ha nel verificarsi di un certo evento, il quale dipende dalle informazioni che ha riguardo l'evento e la.

Non si può definire se un approccio aia migliore dell'altro, ma si può notare che entrambe le interpretazioni soddisfano gli assiomi della probabilità:
Assumendo che Ω sia l'insieme di casi possibili, che $|\Omega| = n$ sia la sua cardinalità e che E_i, con i=1,2..,n siano gli eventi di Ω, allora:

$$0 \leq P(E_i) \leq 1 \tag{3}$$

$$P(\Omega) = 1 \tag{4}$$

$$P(\cup_{n \to \infty}^{\infty} E_i) = \Sigma_{n \to \infty}^{\infty} P(E_i) \tag{5}$$

La differenza sostanziale dei due approcci si deve cercare nel concetto che viene assegnato alla probabilità, che inciderà sull'utilizzo e l'importanza del teorema di Bayes, anch'esso comune nelle due interpretazioni.

In questa sezione verrà analizzata in maniera più approfondita la probabilità e l'inferenza Bayesiana. Per cominciare verranno richiamate alcune definizioni che sono alla base di questa, quali il teorema di Bayes, probabilità condizionata, distribuzione a priori e a posteriori.

0.2 Richiamo definizioni

0.2.1 Distribuzione

La funzione di distribuzione descrive la probabilità che un evento assuma un particolare valore.

0.2.2 Probabilità condizionata

In teoria della probabilità la probabilità condizionata di un evento A rispetto a un evento B è la probabilità che si verifichi A, sapendo che B si è verificato. Viene indicata come:

$$P(A|B) = \frac{P(A,B)}{P(A)} \tag{6}$$

dove $P(B) \neq 0$ e P(A,B) indica la probabilità congiunta.

La probabilità condizionata esprime una "correzione" delle aspettative per A, dettata dall'osservazione di B.

0.2.3 Probabilità congiunta

La probabilità congiunta indica la probabilità che due eventi A e B si verifichino nello stesso momento. Si indica come:

$$P(A,B) = P(A|B)P(A) \tag{7}$$

Nel caso in cui due eventi siano indipendenti, ovvero conoscere lo stato di A non ci fornisce alcuna informazione in più sullo stato di B, la probabilità congiunta equivale a:

$$P(A,B) = P(A)P(B) \tag{8}$$

0.2.4 Marginalizzazione

Conoscendo la distribuzione congiunta P(A,B) è possibile ricavare la distribuzione di P(X) usando la marginalizzazione

$$P(A) = \Sigma_B P(A,B)) \tag{9}$$

0.2.5 Teorema di Bayes

Il teorema di Bayes mette in relazione la probabilità condizionata con quella congiunta. Enunciato: La probabilità a posteriori è uguale alla probabilità a priori per likelihood ratio.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}) \tag{10}$$

dove:

- probabilità a priori: corrisponde a P(A), ovvero alla probabilità che l'evento A si verifichi prima di prendere in considerazione le osservazioni.
- probabilità a posteriori: corrisponde a $P(A|B)$, ovvero alla probabilità che l'evento si verifichi avendo considerato le osservazioni. Può essere considerata un aggiustamento della probabilità a priori.
- funzione di verosomiglianza (likelihood): corrisponde a $P(B|A)P(A)$,ovvero alla probabilità che si verifichi un evento A, avendo osservato B.

0.3 Probabilistic reasoning

L'inferenza è il processo che permette partendo dall'osservazione di fatti o circostanze di trarre una conclusione. Il processo di inferenza viene fatto costantemente dalla mente umana in situazioni di incertezza, per esempio:
"Attraversando la strada tutti abbiamo fatto ciò: Sei di fretta, quindi invece di arrivare alle prossime strisce pedonali guardi a destra e a sinistra e vedi una macchina ad una distanza sufficiente perchè tu riesca ad attraversare la strada senza essere investito. Quindi tu cammini (e immagino) attraversi la strada illeso.[Las12]

In questo esempio non si sa con certezza che la macchina vista non stesse andando troppo veloce da investirci eppure, nella maggior parte dei casi, si riesce a fare la scelta giusta.

E' possibile eseguire questo tipo processo anche in campo statistico, e si parla di Probabilistic Reasoning, anche questo è soggetto ai due tipi di interpretazione.

0.3.1 Bayesian reasoning

Nella prospettiva bayesiana la probabilità descrive l'incertezza con il quale è possibile prevedere l'esito del fenomeno in questione. Questa incertezza può derivare dall'intrinseca variabilità del processo, ma anche dalla nostra imperfetta conoscenza di esso. E' vista come un'estensione della logica proposizionale. "L'idea base è quella di fare inferenza (bayesiana), ossia sviluppare delle previsioni (probabilità a posteriori), su quantità non note in base all'evidenza disponibile e a un'ipotesi iniziale. Quest'ultima è la cosiddetta probabilità a priori, che costituisce l'elemento di soggettività delle analisi bayesiane. "E' possibile fare tutto ciò utilizzando il teorema di Bayes che viene interpretato in questo modo:

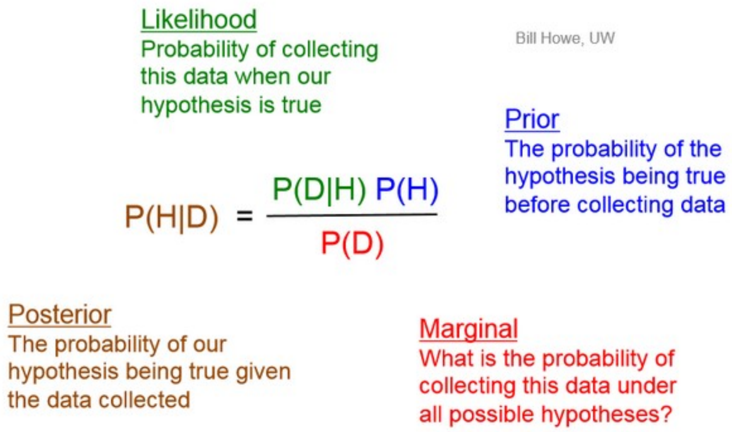

Figura 1: Bayes theorem.Fonte:

0.3.2 Analisi Bayesiana

L'analisi Bayesiana consiste quindi nel:

- creare un modello di probabilità che rappresenti i dati
- identificare le evidenze
- definire le probabilità a priori
- eseguire l'inferenza

Esempio 1:

Un giorno Tracey si sveglia è realizza che l'erba del suo giardino è bagnata. Avrà piovuto durante la notte oppure si è dimenticata di spegnere l'irrigatore? Dopodichè si rende conto che anche l'erba del suo vicino è bagnata. Qual è la probabilità che Tracey abbia lasciato l'irrigatore acceso durante la notte?

Modello:

$R \in 0, 1 \rightarrow R = 1$ ha piovuto, 0 altrimenti
$S \in 0, 1 \rightarrow S = 1$ Tracey ha lasciato l'irrigatore acceso, 0 altrimenti
$J \in 0, 1 \rightarrow J = 1$ l'erba di Jack *è* bagnata, 0 altrimenti
$T \in 0, 1 \rightarrow T = 1$ l'erba di Tracey *è* bagnata, 0 altrimenti
E' necessario inoltre definire il tipo di dipendenza che c'è tra i diversi eventi, queste relazioni possono essere rappresentate in modo grafico attraverso una Bayes Network, che verrà spiegata approfonditamente successivamente: $P(T|S, R)$ e $P(J|R)$

Dal problema evince che T=1 e R=1. Supponiamo che le probabilità a priori siano: P(R)=0.2 P(S)=0.1 Mentre le altre probabilità corrispondono a:

$$p(J = 1|R = 1) = 1, p(J = 1|R = 0) = 0.2$$

$$p(T = 1|R = 1, S = 0) = 1, p(T = 1|R = 1, S = 1) = 1$$

$$p(T = 1|R = 0, S = 1) = 0.9, p(T = 1|R = 0, S = 0) = 0$$

.

A questo punto si può iniziare ad eseguire l'inferenza andando ad applicare il teorema di Bayes. L'obiettivo è calcolare P(S—T,J). Iniziamo con il calcolare P(S—T) e vediamo come cambiano le due probabilità con l'aggiunta di un evidenza.

$$P(S|T) = \frac{P(S,T)}{P(T)} = \frac{\Sigma_{J,R}\, P(T,J,R,S)}{\Sigma_{J,R\ ,S}\, P(T,J,R,S)} \tag{11}$$

$$= \frac{\Sigma_{J,R}\, P(J|R)P(T|R,S)P(R)P(S)}{\Sigma_{J,R\ ,S}\, P(J|R)P(T|R,S)P(S)} \tag{12}$$

$$= \frac{\Sigma_R P(T|R,S)P(R)P(S)}{\Sigma_{R,S}\, P(T|R,S)P(S)P(R)} = 0,3340 \tag{13}$$

Si può notare che la probabilità a posteriori è aumentata rispetto a quella a priori, tenendo conto dell'evidenza che T = 1.

Ora andiamo a calcolare $P(S|T,J)$ e vediamo come si comporta la probabilità a posteriori in questo caso.

$$P(S|T,J) = \frac{(S,T,J)}{P(T,J)} = \frac{\Sigma_R P(T,J,R,S)}{\Sigma_{R,S}\, p(T,J,R,S)} \tag{14}$$

$$= \frac{\Sigma_R P(J|R)P(T|R,S)P(R)P(S)}{\Sigma_R P(J|R)P(T|R,S)P(R)P(S)} = 0.160 \tag{15}$$

La probabilità che Tracey abbia lasciato l'irrigatore acceso, data l'evidenza che anche l'erba di Jack è bagnata è inferiore di quella che tiene conto solo dell'erba di Tracey. Questo perchè l'osservazione che l'erba di Jack è bagnata aumenta la probabilità che sia vero che abbia piovuto.

0.4 Belief Network

Definizione:
Una belief Network è una distribuzione della forma:

$$P(x_1, ..., x_n) = \Pi_{i=1}^{D}(x_i|Pa(x_1)) \tag{16}$$

dove $Pa(x_i)$ equivale alle variabili parenti di x_i.

Come è stato accennato precedentemente è possibile rappresentare un modello probabilistico attraverso una Belief Network(Bayesian Belief Network) che non è altro che un DAG (Direct Acyclic Graph) che è in grado di rappresentare le dipendenze condizionali di un modello.

All'interno della Belief network ogni nodo rappresenta una variabile ed è associato ad una funzione di probabilità,che prende in input un insieme di possibili valore che potrebbe assumere il nodo padre e restituisce la distribuzione di probabilità del nodo. Gli archi, invece, rappresentano le dipendenze condizionali delle variabili. L'assenza di un cammino tra due nodi implica che le due variabili sono condizionalmente indipendenti. Rappresentare la dipendenza tuttavia non è così immediato, infatti bisognerebbe tenere conto di alcuni pattern, quali:

- $A \to B \to C$: viene chiamato "chain ". A primo impatto si potrebbe pensare che il valore di A influenzi C attraverso B, ma non è così. Infatti, è osservabile che condizionando la varibile B =1 il valore di C non varierà indipendentemente dal valore che assume B. Quindi le variabili A e C sono indipendenti.

- $A \leftarrow B \to C$: viene chiamato "fork". Anche in questo caso abbiamo che A e C sono indipendenti, nonostante sembrerebbero essere correlati attraverso B.

- $A \to B \leftarrow C$: viene chiamato "collider". A e C sono indipendenti in linea generale, ma condizionando B diventano dipendenti. Più precisamente si ha che A e B sono inversamente proporzionali, quindi un valore alto di A spiega il perchè di un valore basso di C.

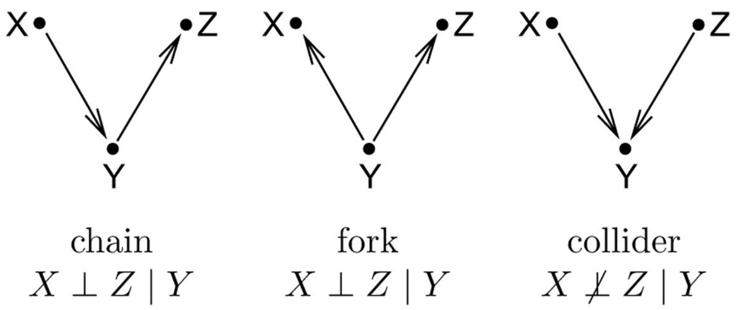

Figura 2: Fonte: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-S6-S5/figures/3

Inoltre, su una Bayesian Network è anche possibile fare inferenza, sfruttando la probabilità condizionale, che è gia definita ,data la definizione di belief network: A partire da questa viene fatta la valutazione della variabile desiderata. Per fare inferenza, tutta via si possono sfruttare anche altre tecniche, come ad esempio MCMC, che si presta bene ad inferire su reti di grandi dimensioni.

Esempio 2:

Si andrà a rappresentare una belief network per l'esempio 1.

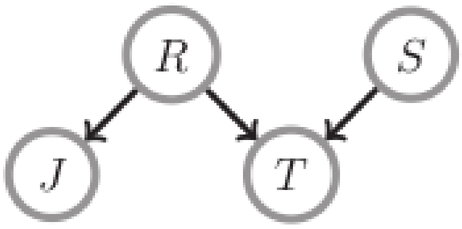

Figura 3:

E' possibile vedere che sono presenti tutte le varibili del problema e che si è riusciti a modellare le dipendenze, infatti tra R e S non sono presenti dei cammini e J è influezato solo da R, mentre T dipende da R e S.