Inferenza in ambito fuzzy: il ruolo della probabilità condizionata coerente

Giulianella Coletti[a]

[a]University of Perugia

9 giugno 2015

Cosa significa fare inferenza ?

Quando una procedura inferenziale può dirsi soddisfacente?

Fare inferenza è il processo che permette di extendere a nuovi eventi (condizionati) l'informazione relativa ad una classe di eventi (condizionati), sintetizzata da una assegnazione numerica o qualitativa, compatible con una measura di incertezza.

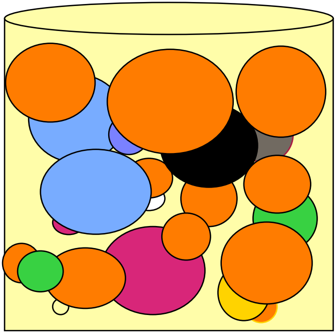
Ogni buona procedura inferentiale deve:

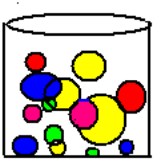
- usare tutta l'informatione disponibile;
- non introdurre durante il processo informazione fittizia o interventi arbitrari dell'esperto o del decisore;
- mantenere la coerenza globale con il framework di riferimento (che nonè necessariamente quello iniziale).

Qualche volta in uno stesso prblema l'incertezza è trattata con framework diversi per esempio quello probabilistico e quello degli insiemi fuzzy, poiché l'informazione non è completa per cause diverse.

ESEMPIO: eventi booleani incerti o eventi fuzzzy perché definiti utilizzando il linguaggio naturale.

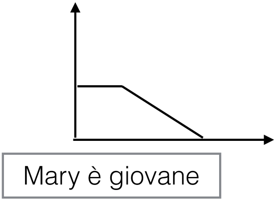
Quale è la probabilità di estrarre una palla grande da un'urna contenente palle di diversi diametri $\{d_1, ..., d_n\}$?

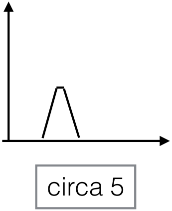


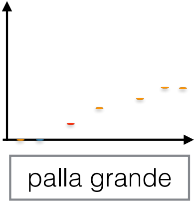


Quale è la probabilità di estrarre una palla chiara e grande supposto che la maggior parte sono chiare e poche sono piccole?

Un sottoinsieme fuzzzy di uno spazio S è individuato da una funzione di appartenenza $\mu(s)$, a valori in $[0, 1]$.


Mary è giovane


circa 5


palla grande

OPERAZIONI: generalizzano quelle insiemistiche classiche utilizzando t-norme e t-conorme e funzione di complementazione.

ESEMPI:

$(\max, \min, 1-x)$

$(\oplus_L, \odot_L, 1-x)$

$(x+y-xy, xy, 1-x)$

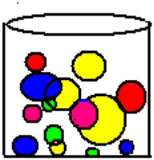
casi particolari classe di Frank:

$$x \oplus y = x + y - y \odot y$$

Molte altre t-norme e t-conorme e complementazione sono state proposte

Importante: se A è un sottoinsieme fuzzy dello spazio S

$$A \cup A^c \neq S$$



Quale è la probabilità di estrarre una palla chiara e grande supposto che la maggior parte sono chiare e poche sono piccole?

Due vie possibili:

• Definire il concetto di probabilità in questo nuovo ambito (logica non booleana per il dominio)
(Mundici, Di Nola.....relativamente alle operazioni/logica di Lukasiewicz)

• Oppure..... dare una definizione (interpretazione) di fyzzy set utilizzando il framework probabilistico.

An interpretation

Interpretatione (C.- S. 2002) della membership di un sottoinsieme fuzzy di uno spazio $\mathcal{C}_X$ in termini di assegnazione coerente di probabilità condizionata, definita in un insieme di eventi condizionati $E|H_i$ con $\{H_i\}$ partizione (verosimiglianza dal punto di vista sintattico).

Definizione

Dato un insieme $\mathcal{C} = \mathcal{G} \times \mathcal{B}^o$ *di eventi condizionati, with* $\mathcal{G}$ *algebra booleana,* $\mathcal{B} \subseteq \mathcal{G}$ *insieme additivo, e* $\mathcal{B}^o = \mathcal{B} \setminus \{\emptyset\}$ *, una funzione* $P : \mathcal{C} \to [0,1]$ *è una probabilità condizionata se:*

(i) $P(H|H) = 1$, *per ogni* $H \in \mathcal{B}^o$ *,*

(ii) $P(\cdot|H)$ *è una probabilità (finitamente additiva) in* $\mathcal{G}$ *per ogni* $H \in \mathcal{B}^o$ *,*

(iii) $P\big((E \wedge A)|H\big) = P(E|H) \cdot P\big(A|(E \wedge H)\big)$, *per ogni* E, $A \in \mathcal{G}$ *and* E, $E \wedge H \in \mathcal{B}^o$.

Definizione

Una assegnazione $P(\cdot|\cdot)$ in un insieme arbitrario di eventi condizionati $\mathcal{C} = \{E_i|H_i\}$ è consistente *con una probabilità condizionata se esiste una probabilità conditionata P' in $\mathcal{C}' = \mathcal{G} \times \mathcal{B}^o$ (con $\mathcal{G}$ Boolean algebra generata dagli eventi $\{E_i, H_i\}$ e $\mathcal{B}^o$ insieme additivo, generato dagli eventi $\{H_i\}$), che estende P.*

Caratterizationi:

• in termini di coerenza di scommesse condizionate (de Finetti, Williams, Regazzini, Holzer,...)

• in termini di minimizzazione della penalizzazione (Gilio)

• in termini di esistenza (per ogni sottoinsieme finito) di una classe di probabilità (non conditionate) che vanno d'accordo con l'assegnazione data e che sono le soluzioni di una sequenza di sistemi lineari (C, C.-S., Gilio).

Risultato fondamentale (essenzialmente dovuto a de Finetti):

Una assegnazione coerente in un insieme $\mathcal{K}$ di eventi condizionati può essere estesa in modo coerente ad ogni soprainsieme di $\mathcal{K}$

Z una variabile (non necessariamente numerica), con codominio $\mathcal{C}_Z$,
per ogni $z \in \mathcal{C}_Z$, z indica l'evento $\{Z = z\}$.
φ è una qualunque *proprietà* relativa a Z .
E_φ={Tu dichiari che Z è φ}
$E_\varphi|z$ ={Tu dichiari che Z è $\varphi|\{Z = z\}$}
$P(E_\varphi|z) = \mu_\varphi(z)$

Allora la membership function $\mu_\varphi(z)$ è una misura di quanto probabile ritieni che Tu dichiari che Z è φ, quando Z assume i diversi valori del suo codominio.

Un sottoinsieme fuzzy E^*_φ è la coppia (E_φ, μ_φ)

Quali restrizioni impone questa interpretazione sulla Membership??

NESSUNA

È un insieme fuzzy così definito una effettiva generalizzazione di un insieme (crisp)?

SI

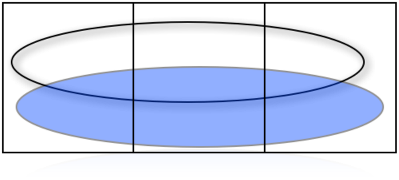
Theorem

(Regazzini 1987, C.-S.2002)
Sia $\mathcal{C}$ *una famiglia di eventi condizionati* $\{E|H_i\}_{i\in I}$*, con card*(I) *arbitraria e gli eventi* H_i *che sono* partizione *di* Ω*.* *Ogni funzione* $f : \mathcal{C} \to [0,1]$ *tale che that* $f(E|H_i) = 0$ *se* $E \wedge H_i = \emptyset$ *e* $f(E|H_i) = 1 \text{ if } H_i \subseteq E$ *è una assegnazione di probabilità coerente.*

Inoltre se la sola assegnazione coerente in $\mathcal{C}$ *prende solo valori in* $\{0,1\}$*, allora si ha* $H_i \wedge E = \emptyset$ *per ogni* $H_i \in \mathcal{H}_o$ *, e è* $H_i \subseteq E$ *per ogni* $H_i \in \mathcal{H}_1$ *, dove* $\mathcal{H}_r = \{H_i : P(E|H_i) = r\}$ *,* $r = 0, 1$ *.*

an interpretation of the membership

Date due proprietà φ, ψ, gli eventi E_φ e E_ψ sono logicamente indipendenti, rispetto alla partizione $\{Z = z\}$,



anche se ψ è una qualunque variazione di φ ottenuta con avverbi accrescitivi o diminutivi o anche se ψ è $\neg\varphi$.

ATTENZIONE:

$$\neg(E_\varphi) \neq E_{\neg\varphi}$$

Quando E_φ e E_ψ sono logicamente indipendenti rispetto alla partizione $\{Z = z\}_{z \in \mathcal{C}_Z}$, l'assegnazione $P(E_\varphi \wedge E_\psi | z)$ è coerente se e solo se vale

$$\begin{aligned}(P(E_\varphi|z) \odot_L P(E_\psi|z)) &\leq P(E_\varphi \wedge E_\psi|z) \\ &\leq \min(P(E_\varphi|z), P(E_\psi|z)),\end{aligned} \tag{1}$$

e tutti i valori nell'intervallo sono coerenti

La coerenza assicura che le operazioni di unione and interseczione dei relativi sottoinsiemi fuzzy $E^*_\varphi,\ E^*_\psi$ possono essere calcolate con una qualunque delle t-norme e t-conorme di Frank e la complementazione con $1-(\cdot)$, mantenedo la coerenza.

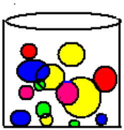
$$P(E_{\varphi_i} \wedge E_{\varphi_j}|z) = P(E_{\varphi_i}|z) \odot_F P(E_{\varphi_j}|z)$$

e

$$P(E_{\varphi_i} \vee E_{\varphi_j}|z) = P(E_{\varphi_i}|z) \oplus_F P(E_{\varphi_j}|z)$$

Per eventi E non logicamente independenti l'ampliamento è regolato dalleregole della coerenza

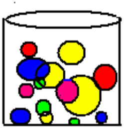
In questo contesto un evento fuzzy coincide con l'evento E_φ e la sua probabilità deve essere calcolata estendendo $\mu_\varphi(z) = P(E_\varphi|z)$ e la distibuzione di probabilità su Z, (utilizzando le regole della probabilità) an interpretation of the membership



La probabilità di estrarre “pallina grande” è la probabilità che tu dichiari grande la pallina estratta

$$P(E_{\varphi}) = \sum_i P(d_i)\mu_{\varphi}(d_i)$$

Questa interpretatione di un fuzzy set come asssegnazione di probabilità condizionata coerente ovviamente fornisce un metodo naturale per scegliere l'elemento più probabile di $\mathcal{C}_Z$ usando sia informazione statistica che informazione fuzzy.



Siamo interessati a trovare quale tra i diametri delle palline sia più probabile che sia stato estratto se si suppone che sia stata estratta una palla grande.

P è una distributione di probabilità sugli elementi di $\mathcal{C}_Z$ e l'informatione fuzzy è expressa da una membership function

$\mu_\varphi(z) = P(E_\varphi|z)$.
Usando la formula di Bayes è immediato calcolare, per ogni $z \in \mathcal{C}_Z$, il valore $P(z|E_\varphi)$ come

$$P(z|E_\varphi) = \frac{P(z)\mu_\varphi(z)}{\sum_z \mu_\varphi(z)P(z)}$$

Osservazione curiosa......

Se abbiamo più di un evento E_φ.....

Theorem

Supponiamo di avere un insieme di eventi $\mathcal{C} = \{E_j|H_i\}$ $(i = 1, ..., n; 1 = i, ..., m)$ *con gli eventi* H_i *che formano una partizione of* Ω *e gli* E_j *logicamente independendenti rispetto alla partizione e sia* $P(H_i)$ *una distributione di probabilità in* $\{H_1, ..., H_m\}$*. Allora per ogni funzione* $P : \mathcal{C} \rightarrow [0, 1]$ *con* $P(E_j|H_i) = 0 \text{ if } E_j \wedge H_i = \emptyset \text{ and } P(E_j|H_i) = 1 \text{ if } H_i \subseteq E_j$*, l'assgnazione globale*

$$\{P(E_j|H_i), P(H_i)\}$$

è coerente

Per ogni t-norma $\odot$ della classe di Frank l'estensione, calcolata via l'integrale di Daniel

$$P_{\odot}(E_{\varphi_i}) = \int \mu_{\varphi_i}(x) dP(x)$$

$$P_{\odot}(E_{\varphi_i} \wedge E_{\varphi_i}) = \int (\mu_{\varphi_i} \odot \mu_{\varphi_j})(x) dP(x)$$

è ancora coherente.

Ovviamente non è l'unica

A questo punto l'estensione agli eventi $E_{\varphi_i} \vee E_{\varphi_j}$ è univocamente determinata dalla coerenza, e soddisfa l'equazione di Frank, con $\oplus$ t-conorma duale di $\odot$:

$$P_{\odot}(E_{\varphi_i} \vee E_{\varphi_j}) = \int (\mu_{\varphi_i} \oplus \mu_{\varphi_j})(x) dP(x) =$$

$$= \int \mu_{\varphi_i}(x) dP(x) + \int \mu_{\varphi_j}(x) dP(x) - \int (\mu_{\varphi_i} \odot \mu_{\varphi_j})(x) dP(x).$$

Questa assegnazione di probabilità condizionata coerente $P_{\odot}$ può essere ulteriormente estesa ad eventi condizionati $A|B$ dove A, B sono eventi dell'algebra generata $\mathcal{B}$, con $B \neq \emptyset$.

Questa estensione in generale nonè unica, ma lo è per gli eventi $A = E_{\varphi_i}$ e $B = E_{\varphi_j}$ $(i \neq j)$, con $P_{\odot}(E_{\varphi_j}) > 0$ per i quali necessariamente si ha:

$$P_{\odot}(E_{\varphi_i}|E_{\varphi_j}) = \frac{\displaystyle\int (\mu_{\varphi_i} \odot \mu_{\varphi_j})(x)dP(x)}{\displaystyle\int \mu_{\varphi_j}(x)dP(x)}.$$

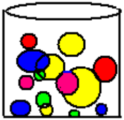
Con queste poche cose siamo in grado di affrontare problemi come i seguenti, sottopostici da Zadeh:

- If we do not know which urn has been chosen, what is the probability that a ball drawn at random is small, under the hypothesis that in the urn “most balls are large”?

- Usually it takes George about an hour to get home from work. Usually George leaves office at about 5pm. What is the probability that George is home at 6.15pm.

Finora abbiamo supposto che fosse disponibile la distribuzione di probabilità (o la densità) delle variabili di riferimento, ma in molte situazioni questo non è vero.
Per esempio per valutare l'incertezza negli elementi di un database disponiamo di informazioni statistiche relative a qualche attributo e di membrship relative a proprietà di altri attributi
Prima di tutto ' necessario testare la coerenza globale delle assegnazioni, e, se questa vale, possiamo estendere questa assegnazione ad altri eventi.

UN CASO DIVERSO Non conosciamo la composizione dell'urna rispetto aidiametri, ma conosciamola probbilità di altri eventi collegati, quali per esempio:



$B_1 =$"il diametro è minore di x",
$B_2 =$ " il diametro è maggiore di y",
$B_3 =$"il diametro ha una lunghezza compresa nell'intervallo $[z, z']$".