Probabilitá condizionata

12 marzo 2017

Si introduce il concetto di indipendenza stocastica e la regola del prodotto per eventi indipendenti. Si discute di come un processo con memoria introduca un condizionamento fra eventi e ne si mostra l'effetto su una rappresentazione ad albero dello spazio degli stati. Si considera in seguito la relazione tra la correlazione fra eventi e il condizionamento. Infine, si definisce la probabilitá condizionata e, su tale base, la regola del prodotto generalizzata.

1 *Indipendenza Stocastica*

Consideriamo l'esperimento $\mathcal{E}$ ="Lancio di due monete " (monete bilanciate) e l'evento

$$E = \text{"Esce almeno una volta testa".}$$

É possibile rappresentare lo spazio degli stati dell'esperimento mediante una tabella che mostra tutti i possibili esiti.

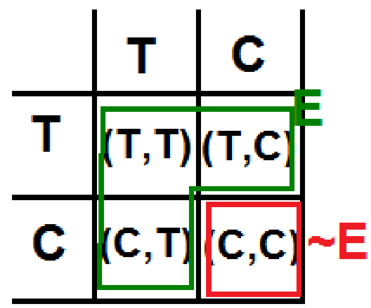


Figura 1: Tabella degli esiti

Lo spazio di campionamento dell'esperimento, come si nota dalla tabella 1 ha cardinalitá $\sharp(S) = 4$ ovvero tutte le combinazioni possibili dei lanci mentre la cardinalitá dell'evento é $\sharp(E) = 3$ escludendo il caso in cui le due monete hanno entrambe dato "Croce".

Indichiamo con $M_i = T$ oppure $M_i = C$ l'esito elementare per il lancio della monetaM_i, $i = 1, 2$. Possiamo allora scrivere E come l'esito che si ottiene quando una della due monete produce "Testa".

$$E = \{M_1 = T\} \text{ OR } \{M_2 = T\}$$

La probabilitá dell'evento E si calcola dunque come

$$\begin{aligned}
P(E) &= P(\{M_1 = T\} \text{ OR } \{M_2 = T\}) \\
&= P(M_1 = T \cup M_2 = T) \\
&= P(M_1 = T) + P(M_2 = T) - P(M_1 = T \cap M_2 = T) \\
&= \frac{1}{2} + \frac{1}{2} - \left[\frac{1}{2} \times \frac{1}{2}\right] \\
&= \frac{3}{4}
\end{aligned}$$

Si noti che abbiamo assunto intuitivamente:

$$P(M_1 = T \cap M_2 = T) = P(M_1 = T)P(M_2 = T) = \frac{1}{2} \times \frac{1}{2},$$

ovvero, gli esiti del lancio delle due monete non dipendono uno dall'altro: sono stocasticamente indipendenti. La definizione di Indipendenza Stocastica é dunque

Definizione 1.1 (Indipendenza Stocastica) *Dati due eventi A e B, se $P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B)$, allora si dice che gli eventi A e B sono **stocasticamente indipendenti**(s.i.).*

Osservazione 1.2 *Per calcolare la probabilitá dell'evento E, si puó evitare di usare una rappresentazione tabellare e adottare un approccio inverso, ovvero determinando la probabilitá dell'evento complementare $\sim E =$ "Nessuna testa" che equivale a "Tutte (due) croci" per poi sfruttare la regola del complemento: $P(E) = 1 - P(\sim E)$*

$$\begin{aligned} P(E) &= 1 - P(M_1 = C \cap M_2 = C) \\ &= 1 - P(M_1 = C)P(M_2 = C) \\ &= 1 - \left[\frac{1}{2} \times \frac{1}{2}\right] \\ &= \frac{3}{4} \end{aligned}$$

Sotto l'ipotesi di indipendenza stocastica, possiamo realizzare $\mathcal{E}$ in due modi: lanciando contemporaneamente due monete M_1 e M_2, oppure utilizzare la stessa moneta in due lanci consecutivi $M_{t=1}$ e $M_{t=2}$: nel secondo caso l'indice $i = 1, 2$ funge da indice temporale.

Possiamo dare una visualizzazione dinamica e calcolare le probabilitá di interessse é tramite l'**albero di probabilitá** (in inglese, *chance tree*) solitamente utilizzato quando un esperimento é si svolge sequenzialmente in piú stadi (in questo caso gli stadi sono i lanci):

- Per ogni stadio (nodo) ci sono tanti percorsi radice / foglia quanti sono gli esiti elementari
- Il numero totale dei percorsi rappresenta il numero totale di esiti possibili (foglie dell'albero)
- Ad ogni percorso o ramo dell'albero é associata la probabilitá corrispondente all'evento; gli archi sono etichettati con le probabilitá "parziali"

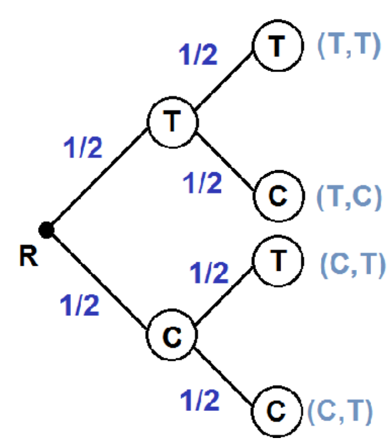


Figura 2: Visualizzazione ad albero di due lanci consecutivi

Osservazione 1.3 *Seguendo i rami del percorso si ottengono tutti gli esiti composti descritti nella tabella precedente (Vedi Figura 1): la probabilitá complessiva si ottiene moltiplicando tra loro tutte le probabilitá incontrate nel percorso per raggiungere l'esito. Si ha che, per indipendenza stocastica:*

$$P(TT) = P(M_1 = T \cap M_2 = T) = P(M_1 = T)P(M_2 = T) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

ecc. Cosí, poiché TT, TC, CT sono disgiunti:

$$P(E) = P(TT) + P(TC) + P(CT) = \frac{3}{4}$$

2 Esperimenti con "memoria" e condizionamento

Possiamo considerare una sequenza di lanci indipendenti come un processo senza memoria: ad ogni lancio la probabilitá di avere testa o croce é sempre pari a $\frac{1}{2}$.

Un tipo di esperimento che puó introdurre memoria ovvero una dipendenza é quello basato sul modello dell'urna (contenente oggetti)

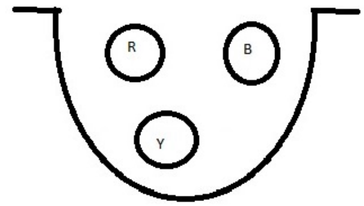


Figura 3: Modello dell'urna

Con questo tipo di modello posso operare in due modalitá:

- con **reimbussolamento**: estraggo l'oggetto e lo rimetto nell'urna;
- senza **reimbussolamento**: estraggo l'oggetto e non lo rimetto nell'urna.

La misura dello spazio di campionamento S corrisponde al numero di oggetti presenti all'interno dell'urna. S viene quindi modificato se estraggo un elemento e non lo reinserisco. L'estrazione senza reimbussolamento introduce dunque una memoria nel processo: l'esito che ottengo al tempo t_{i+1} dipende dall'esito al tempo t_i.

Ad esempio, supponiamo di avere un'urna strutturata come in figura 4: contiene 3 palline di diverso colore: R, Y, B

Consideriamo l'esperimento $\mathcal{E}$: "estrazione di una coppia di palline".

Lo spazio campionario S ha misura pari a tutte le possibili disposizioni con ripetizione $RR, RB, RY, BR, \cdots$, ovvero $\sharp(S) = 3^2 = 9$

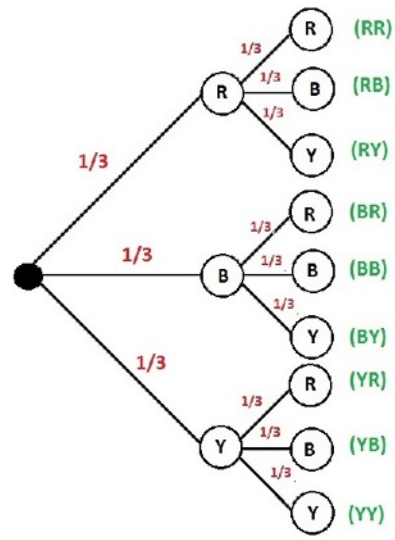


Figura 4: Modello con reimbussolamento

Il numero delle sequenze di k oggetti estratti da n oggetti, ognuno dei quali puó essere ripetuto détta disposizione con ripetizione e si calcola come $D^r_{n,k} = n^k$

Se procediamo con reimbussolamento, possiamo visualizzare S con la rappresentazione ad albero di figura 4.

Poiché, nella modalitá con reimbussolamento, estrazioni successive sono stocasticamente indipendenti troviamo:

$$P(RR) = P(R)P(R) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$$

e cosí per tutti gli altri casi. In termini di misura: $P(RR) = \frac{\sharp(RR)}{\sharp(S)} = \frac{1}{9}$

Senza il reimbussolamento della pallina ottenuta alla prima estrazione, S cambia (ho una pallina in meno) e alla seconda estrazione le probabilitá sono diverse (ho solo piú due possibilitá e non tre), come visualizzato chiaramente nell'albero di figura 5 :

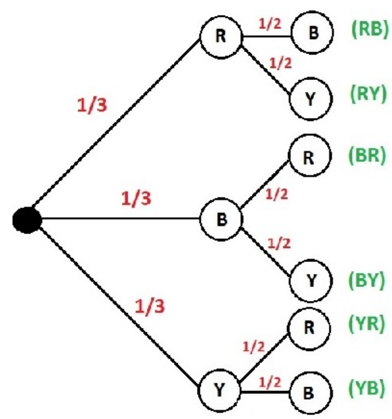


Figura 5: Modello senza reimbussolamento

Lo spazio campionario si é ristretto, infatti ha misura: $\sharp(S) = 3 \times 2 = 6$

In questo caso, per esempio:

$$P(RB) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$$

ovvero il risultato é diverso da quello che si aveva con reimbussolamento, $P(R)P(B) = \frac{1}{3} \times \frac{1}{3}$

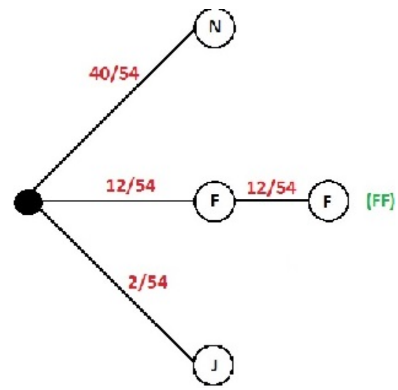
Qui la seconda estrazione é **condizionata** dalla prima: vi é un *effetto* della prima estrazione sulla seconda. Il processo é dunque con memoria.

Un caso analogo é il seguente.

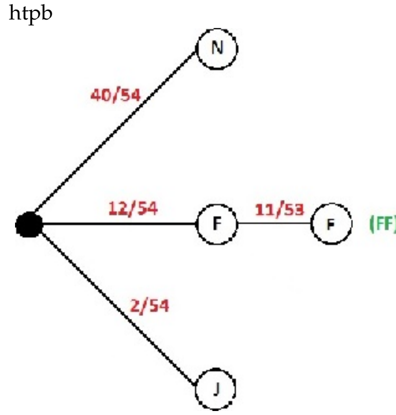
Abbiamo 54 carte suddivise in:

- numeriche: 40;
- figure 12;
- jolly 2.

Consideriamo la possibilitá di pescare due figure di seguito, ovvero consideriamo l'evento



htpb



$$F_1 \text{ AND } F_2 = F_1 \cap F_2$$

Con reimbussolamento si inserisce la carta nel mazzo dopo averla estratta, dunque:

$P(F_1, F_2) = P(F_1 \cap F_2) = P(F1)P(F2) = \frac{12}{54} \times \frac{12}{54} = \frac{144}{2916} = 0.049$

Senza reimbussolamento, non si inserisce la carta nel mazzo dopo averla estratta:

$P(F_1, F_2) = \frac{12}{54} \times \frac{11}{53} = 0.046$

Notiamo che:

- $P(F_1)$ rimane uguale in entrambi i casi, con o senza reimbussolamento;
- nel caso di reimbussolamento $P(F_2) = P(F_1)$;
- nel caso in cui non si reinserisca la figura estratta allora $P(F_2) \neq P(F_1)$.

Nel secondo caso non sto piú calcolando semplicemente $P(F_2)$ ma calcolo piú precisamente

$P(F_2 \texttt{ dopo aver estratto } F_1) =$

$\text{P}(\text{F}_2 \texttt{ dato } F_1) =$

$\text{P}(\text{F}_2 \texttt{ condizionato a } F_1) =$

$\text{P}(\text{F}_2 \mid F_1)$

La notazione $P(\cdot \mid \cdot)$ indica una **probabilitá condizionata**.

Nei due casi, in sintesi:

- con reimbussolamento: $P(F_1 \cap F_2) = P(F_1)P(F_2) = 0.049$
- senza reimbussolamento: $P(F_1 \cap F_2) = P(F_1)P(F_2 \mid F_1) = 0.046$

dove é evidente che $P(F_1)P(F_2) \neq P(F_1)P(F_2 \mid F_1)$

Dal punto di vista strettamente formale, le due procedure di estrazione della coppia $(F1, F2)$ potrebbero essere equivalenti solo se

$P(F_1)P(F_2) = P(F_1)P(F_2 \mid F_1);$
il che implicherebbe
$P(F_2) = P(F_2 \mid F_1),$
ovvero che F_2 non é condizionato, non dipende da F_1: in altri termini F_1, F_2 sarebbero statisticamente indipendenti

In definitiva le due condizioni

1. $P(F_2) = P(F_2 \mid F_1)$
2. $P(F_1 \cap F_2) = P(F_1)P(F_2)$

costituiscono due modi equivalenti per sancire l'indipendenza statistica degli eventi F_1 e F_2

3 *Correlazione statistica e probabilitá condizionata*

Dati due eventi C, B, e calcolate $P(C \mid B)$ e $P(C)$ si danno tre possibilitá:

1. $P(C \mid B) > P(C)$: in questo caso vi é una dipendenza condizionale
2. $P(C \mid B) < P(C)$: anche in questo caso vi é una dipendenza condizionale
3. $P(C \mid B) = P(C)$: in questo caso C, B sono statisticamente indipendenti

Nei primi due casi si puó affermare che vi é **correlazione** statistica fra i due eventi. In particolare nel caso 1) la correlazione é positiva; nel caso 2), correlazione negativa.

Il terzo caso dice semplicemente che i due eventi sono **scorrelati**.

Osservazione 3.1 *In generale, la presenza di una dipendenza probabilistica condizionale per cui $P(B) \neq P(B \mid A)$ puó verificarsi in due situazioni:*

- *vi é una semplice correlazione o dipendenza statistica tra gli eventi A e B: per esempio, $A =$ "vivere in un'area geografica" (dove potrebbe essere stoccato qualche materiale radioattivo o nocivo, un evento latente o incognito che non é modellato né da A né da B, il quale é invece la vera causa patologica), $B =$ "aver contratto un tumore".*

Correlazione vs. causazione

- *vi é una effettiva causazione (per esempio, in senso fisico): per esempio, $A =$ "aver maneggiato materiale radioattivo", $B =$ "aver contratto un tumore"*

Non é detto che la correlazione implichi una causazione, invece la causazione induce necessariamente una correlazione.

P

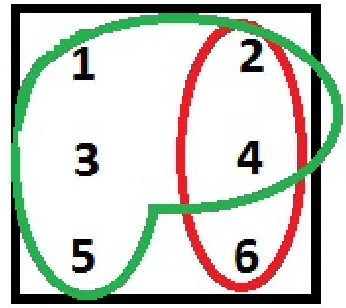


Figura 6: evento $B =$ "numero" < 6

Valutiamo ora la correlazione nei seguenti casi in un esperimento costruito nel seguente modo.

Si lancia un dado. L'evento di interesse é: $C =$ "numero pari". Per tale evento, a priori sappiamo che

$$\sharp(S) = 6$$

$$\sharp(C) = 3$$

$$P(C) = \frac{\sharp(C)}{\sharp(S)} = \frac{1}{2}$$

Quando il dado viene tirato, non vedo direttamente l'esito. C'é peró un osservatore che mi fornisce un'informazione che puó essere correlata o no al risultato. Tale informazione costituisce l'evento B e la probabilitá di C é condizionata a B.

Consideriamo i seguenti casi

1. Informazione: evento $B =$ "numero" < 6". Lo spazio di campionamento si riduce a $1, 2, 3, 4, 5$ diventando:

$$\sharp(S) = 6 \to \sharp(S \mid B) = 5$$

. Anche la misura di C si riduce (l'esito numero pari 6 é ormai escluso secondo l'informazione fornita) :

$$\sharp(C \mid B) = 2$$

Quindi, se devo determinare la probabilitá di C condizionata a questa informazione (cioé la probabilita dell'evento $C \mid B$), intuitivamente:
$P(C \mid B) = \frac{\sharp(C|B)}{\sharp(S|B)} = \frac{2}{5} = 0.4$
Notiamo che $P(C \mid B) < P(C)$ Questo é un caso dove si ha riduzione: siamo in presenza di una correlazione negativa.

2. Informazione: evento $B =$ "numero" > 1"

Ripercorrendo il ragionamento fatto sopra:

$$\sharp(S \mid B) = 5$$

$$\sharp(C \mid B) = 3$$

$P(C \mid B) = \frac{\sharp(C|B)}{\sharp(S|B)} = \frac{3}{5} = 0.6$
In questo caso la situazione si é ribaltata: $P(C \mid B) > P(C)$ Si parla di correlazione positiva

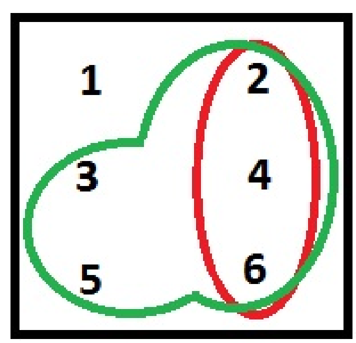


Figura 7: evento $B =$ "numero" > 1"

3. Informazione: evento $B =$ "numero" < 5". Qui $S \mid B = \{1, 2, 3, 4\}$

Pertanto:

$$P(C \mid B) = \frac{\sharp(C \mid B)}{\sharp(S \mid B)} = \frac{2}{4} = \frac{1}{2} = 0.5$$

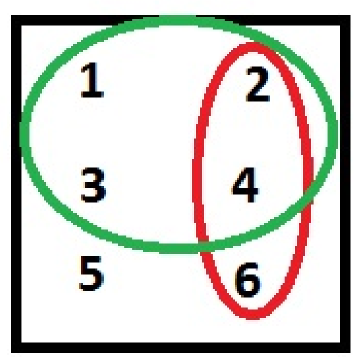


Figura 8: evento $B =$ "numero" < 5"

Quindi $P(C \mid B) = P(C)$ gli eventi sono scorrelati .

Notiamo che questo accade perché $C \mid B$ ha la stessa proporzione in $S \mid B$ di quella che C aveva in S (cioé $\frac{1}{2}$)

Facendo riferimento alle figure, si nota facilmente che le restrizioni imposte dall'informazione B si possono scrivere come

$$(S) \to (S \mid B) = S \cap B = B$$

$$(C) \to (C \mid B) = C \cap B$$

Si noti inoltre che il nostro calcolo "intuitivo" della probabilitá $P(C \mid B)$ puó essere esplicitato come:

$$P(C \mid B) = \frac{\sharp(C \cap B)}{\sharp(B)} = \frac{\frac{\sharp(C \cap B)}{\sharp(S)}}{\frac{\sharp(B)}{\sharp(S)}}$$

dove per definizione:

$$\frac{\sharp(C \cap B)}{\sharp(S)} = P(C \cap B)$$
$$\frac{\sharp(B)}{\sharp(S)} = P(B)$$

E' corretto pertanto dare la seguente:

Definizione 3.2 (Probabilitá condizionata) *Dato lo spazio di probabilitá $(S, \mathcal{F}, P)$ considerati due eventi $C, B \in \mathcal{F}$, con $P(B) > 0$,*

$$P(C \mid B) = \frac{P(C \cap B)}{P(B)}$$

é la probabilitá di C condizionata a B.

Mettiamo in evidenza due casi estremi:

1. B esclude C: ovvero $B \cap C = \emptyset$, allora $P(C \mid B) = 0$
2. B implica C: ovvero $B \subseteq C$, allora $P(C \mid B) = 1$

Osservazione 3.3 *La definizione data:*

$$P(C \mid B) = \frac{P(C \cap B)}{P(B)}$$

e quella di una nuova funzione

$$P_B : F \in \mathcal{F} \to [0, 1]$$

che é una misura di probabilitá, ovvero é una funzione che soddisfa gli assiomi di Kolmogorov:

1. $P_B(C \mid B) \geq 0$

2. *per* $C = C_1 \cup C_2 .... \cup C_n$ *con* $C_i \cap C_j = \emptyset$,

$$P_B(C \mid B) = P_B(C_1 \mid B) + P_B(C_2 \mid B) + ... + P_B(C_n \mid B)$$

3. $P_B(B \mid B) = 1$

4 *Generalizzazione della regola del prodotto*

Nel caso particolare di eventi indipendenti, valeva la seguente regola del prodotto

$$P(C, B) = P(C)P(B)$$

A partire dalla definizione di probabilitá condizionata possiamo immediatamente ricavare la seguente:

Prop. 4.1 (Regola generale del prodotto) *Dati due eventi* B, C, *allora*

$$P(B, C) = P(B)P(C \mid B) \tag{1}$$

Osservazione 4.2 *Si noti che la regola per eventi indipendenti é un caso particolare di (1) quando* $P(C \mid B) = P(C)$

Osservazione 4.3 *Si noti che per simmetria:* $P(C, B) = P(C \cap B) = P(B \cap C) = P(B, C)$,

$$P(C)P(B \mid C) = P(B)P(C \mid B)$$

Osservazione 4.4 *Posso generalizzare induttivamente la regola del prodotto a una successione di n eventi:*

$$P(A_1, A_2 ...., A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1) \cdots P(A_n|A_{n-1}, \cdots A_2, A_1)$$