

# Corso di Statistica Facoltà di Economia

a.a. 2010-2011

**francesco mola**

## Lezione n° 12

## Sommario

- Una misura della probabilità
- Il *postulato empirico del caso*
- Schema logico per le applicazioni
- L'approccio bayesiano:  
Il *Teorema di Bayes*

### Una misura della Probabilità

- Data una **Prova** che genera **k eventi** elementari
- $E_1, \dots, E_k$  necessari ,
- $E_1 \cup E_2 \cup \dots \cup E_k = I$ ,
- incompatibili a due a due  
 $E_i \cap E_j = \emptyset$  per ogni  $i \neq j$

ed equiprobabili

$$P(E_1) = P(E_2) = \dots = P(E_k)$$

### Una misura della Probabilità

- Dai postulati si deduce univocamente la misura della probabilità per ciascuno di essi

$$P(E_i) = \frac{1}{k}$$

per ogni  $i = 1, 2, \dots, k$

## Il postulato empirico del caso

- In una successione di prove, ripetute molte volte nelle stesse condizioni, ogni evento si presenta con una frequenza relativa quasi uguale alla sua probabilità;

la differenza tra frequenza relativa e probabilità di un evento tende ad annullarsi all'aumentare delle prove.

## osservazioni

la frequenza è un concetto **a posteriori**, cioè si calcola *dopo* aver compiuto l'esperimento, mentre la probabilità è un concetto **a priori**, cioè si calcola *prima* dell'esperimento e senza che sia necessario effettuarlo;

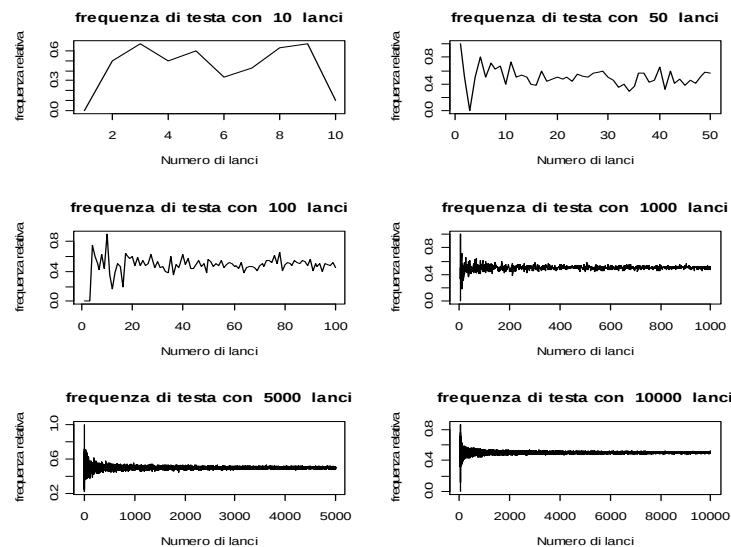
## osservazioni

- nella teoria della probabilità alla parola "caso" non si dà il significato che gli affida il linguaggio comune; per lo statistico, in un esperimento concreto, "**caso**" è l'insieme di quei fattori che egli non ritiene preponderanti nel determinare il risultato della prova;

## osservazioni

- la "tendenza" della frequenza relativa verso la probabilità di un evento *non deve* essere interpretata nel senso dell'analisi matematica (cioè come il limite di una successione); essa scaturisce solo da una universale convinzione circa il comportamento degli eventi casuali.

### Postulato empirico del caso: esempio lancio di una moneta



### Schema logico per le applicazioni

- 1. Individuare correttamente la prova, quindi gli eventi elementari verificando se sono necessari, incompatibili ed equiprobabili in modo da applicare la "regola" trovata precedentemente.

### Schema logico per le applicazioni

- 2. Distinguere correttamente se gli eventi elementari consistono in un solo elemento, o, invece, consistono in più elementi, ed indicarli con simboli di agevole interpretazione.

### Schema logico per le applicazioni

- 3. Esplicitare gli eventi complessi di cui si vuole calcolare la probabilità come unione, intersezione, o negazione degli eventi elementari sopra individuati.

## Schema logico per le applicazioni

- 4. Per l'unione di più eventi, chiedersi se sono incompatibili ricordando che:

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

$$A \cap B \neq \emptyset \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

E le loro ovvie generalizzazioni per più di due eventi.

## Schema logico per le applicazioni

- 5. Per l'intersezione di due eventi, chiedersi se sono indipendenti ricordando che:

$$A \text{ e } B \text{ indipendenti} \Rightarrow P(A \cap B) = P(A) \cdot P(B)$$

$$A \text{ e } B \text{ non indipendenti} \Rightarrow P(A \cap B) = P(A) \cdot P(B | A)$$

### Calcolo delle probabilità partendo da una tabella di dati osservati

Supponiamo di avere un numero elevato di osservazioni riassunte in una tabella

| età   | reddito |             |        |
|-------|---------|-------------|--------|
|       | <25000  | 25000-50000 | >50000 |
| <30   | 5%      | 12%         | 10%    |
| 30-50 | 14%     | 22%         | 16%    |
| >50   | 8%      | 10%         | 3%     |

Possiamo considerare le frequenze relative come un'approssimazione delle probabilità.

### Calcolo delle probabilità partendo da una tabella di dati osservati

Consideriamo i totali di riga e di colonna ed esprimiamo le percentuali in frequenze relative

| età   | reddito |             |        |      |
|-------|---------|-------------|--------|------|
|       | <25000  | 25000-50000 | >50000 |      |
| <30   | 5%      | 12%         | 10%    | 27%  |
| 30-50 | 14%     | 22%         | 16%    | 52%  |
| >50   | 8%      | 10%         | 3%     | 21%  |
|       | 27%     | 44%         | 29%    | 100% |

| età   | reddito |             |        |      |
|-------|---------|-------------|--------|------|
|       | <25000  | 25000-50000 | >50000 |      |
| <30   | 0.05    | 0.12        | 0.10   | 0.27 |
| 30-50 | 0.14    | 0.22        | 0.16   | 0.52 |
| >50   | 0.08    | 0.10        | 0.03   | 0.21 |
|       | 0.27    | 0.44        | 0.29   | 1.00 |

Possiamo considerare le frequenze relative come un'approssimazione delle probabilità.

definisco

- $A$  = un individuo ha reddito maggiore di 50000
- $B$  = un individuo ha + di 30 anni

mi chiedo :

$P(A)$ ?

$P(B)$ ?

$P(A \cap B)$ ?

$P(A \cup B)$ ?

$P(B | A)$ ?

## Definiamo tutti i possibili eventi

$E_1$  = individuo minore di 30 anni e con reddito inferiore a 25000

$E_2$  = individuo tra 30 e 50 anni e con reddito inferiore a 25000

$E_3$  = individuo > 50 anni e con reddito inferiore a 25000

$E_4$  = individuo minore di 30 anni e con reddito tra 25000 e 50000

$E_5$  = individuo tra 30 e 50 anni e con reddito tra 25000 e 50000

$E_6$  = individuo > di 50 anni e con reddito tra 25000 e 50000

$E_7$  = individuo minore di 30 anni e con reddito > 50000

$E_8$  = individuo tra 30 e 50 anni e con reddito > 50000

$E_9$  = individuo > di 50 anni e con reddito > 50000

Considero come probabilità le frequenze relative

$P(E_1) = 0.05$  frequenza relativa di  
individui con età minore di 30 anni  
e con reddito inferiore a 25000

$P(E_2) = 0.14$  frequenza relativa di  
individui tra 30 e 50 anni  
e con reddito inferiore a 25000

$P(E_3) = 0.08$  frequenza relativa di  
individui con età maggiore di 50 anni  
e con reddito inferiore a 25000

$P(E_4) = 0.12$  frequenza relativa di  
individui con età minore di 30 anni  
e con reddito tra 25000 e 50000

$P(E_5) = 0.22$  frequenza relativa di  
individui tra 30 e 50 anni  
e con reddito tra 25000 e 50000

$P(E_6) = 0.10$  frequenza relativa di  
individui con età maggiore di 50 anni  
e con reddito tra 25000 e 50000

$P(E_7) = 0.10$  frequenza relativa di  
individui con età minore di 30 anni  
e con reddito maggiore di 50000

$P(E_8) = 0.16$  frequenza relativa di  
individui tra 30 e 50 anni  
e con reddito maggiore di 50000

$P(E_9) = 0.03$  frequenza relativa di  
individui con età maggiore di 50 anni  
e con reddito maggiore di 50000

Calcolo delle probabilità partendo da una tabella di dati osservati

Come derivare dalla tabella  $P(A)$ ?

| età   | reddito |             |        |      |
|-------|---------|-------------|--------|------|
|       | <25000  | 25000-50000 | >50000 |      |
| <30   | 0.05    | 0.12        | 0.10   | 0.27 |
| 30-50 | 0.14    | 0.22        | 0.16   | 0.52 |
| >50   | 0.08    | 0.10        | 0.03   | 0.21 |
|       | 0.27    | 0.44        | 0.29   | 1.00 |

La somma delle celle in rosso è proprio la risposta

$$P(A) = 0.29$$

Calcolo delle probabilità partendo da una tabella di dati osservati

Come derivare dalla tabella  $P(B)$ ?

| età   | reddito |             |        |      |
|-------|---------|-------------|--------|------|
|       | <25000  | 25000-50000 | >50000 |      |
| <30   | 0.05    | 0.12        | 0.10   | 0.27 |
| 30-50 | 0.14    | 0.22        | 0.16   | 0.52 |
| >50   | 0.08    | 0.10        | 0.03   | 0.21 |
|       | 0.27    | 0.44        | 0.29   | 1.00 |

La somma delle celle in rosso è proprio la risposta

La somma delle celle in arancione ovviamente danno la stessa risposta

$$P(B) = 0.73$$

## Calcolo delle probabilità partendo da una tabella di dati osservati

Come derivare dalla tabella  $P(A \cap B)$ ?

| età   | reddito |             |        |      |
|-------|---------|-------------|--------|------|
|       | <25000  | 25000-50000 | >50000 |      |
| <30   | 0.05    | 0.12        | 0.10   | 0.27 |
| 30-50 | 0.14    | 0.22        | 0.16   | 0.52 |
| >50   | 0.08    | 0.10        | 0.03   | 0.21 |
|       | 0.27    | 0.44        | 0.29   | 1.00 |

La somma delle celle in rosso è proprio la risposta perché sono le celle che soddisfano le condizioni richieste

$$P(A \cap B) = 0.19$$

## Calcolo delle probabilità partendo da una tabella di dati osservati

Come derivare dalla tabella  $P(A \cup B)$ ?

| età   | reddito |             |        |      |
|-------|---------|-------------|--------|------|
|       | <25000  | 25000-50000 | >50000 |      |
| <30   | 0.05    | 0.12        | 0.10   | 0.27 |
| 30-50 | 0.14    | 0.22        | 0.16   | 0.52 |
| >50   | 0.08    | 0.10        | 0.03   | 0.21 |
|       | 0.27    | 0.44        | 0.29   | 1.00 |

La somma delle celle in rosso è proprio la risposta

$$P(A \cup B) = 0.83$$

## Calcolo delle probabilità partendo da una tabella di dati osservati

Come derivare  $P(A \cup B)$  applicando il principio delle probabilità totali?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.29 + 0.73 - 0.19 = 0.83$$

$$P(A \cup B) = 0.83$$

## Calcolo delle probabilità partendo da una tabella di dati osservati

Come derivare dalla tabella  $P(B | A)$ ?

| età   | reddito |             |        |      |
|-------|---------|-------------|--------|------|
|       | <25000  | 25000-50000 | >50000 |      |
| <30   | 0.05    | 0.12        | 0.10   | 0.27 |
| 30-50 | 0.14    | 0.22        | 0.16   | 0.52 |
| >50   | 0.08    | 0.10        | 0.03   | 0.21 |
|       | 0.27    | 0.44        | 0.29   | 1.00 |

La somma delle celle in rosso diviso la cella in arancione è proprio la risposta. E' la probabilità condizionata!

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{0.19}{0.29} = 0.66$$

Calcolo delle probabilità partendo da una tabella di dati osservati

Come derivare  $P(A \cap B)$   
applicando il principio delle probabilità composte?

$$P(A \cap B) = P(A) \cdot P(B | A) = \\ = 0.29 \cdot 0.66 = 0.19$$

Si noti anche che...

$$P(A) = \frac{p(A \cap B)}{P(B | A)} = \frac{0.19}{0.66} = 0.29$$

rieppilogando...

$$P(A) = 0.29$$

$$P(B) = 0.73$$

$$P(A \cup B) = 0.83$$

$$P(A \cap B) = 0.19$$

$$P(B | A) = 0.66$$

### Il teorema di Bayes

Siano dati  $m$  eventi necessari ed incompatibili

$$H_1, H_2, \dots, H_m$$

con  $H_1 \cup H_2 \cup \dots \cup H_m = S$

Necessari

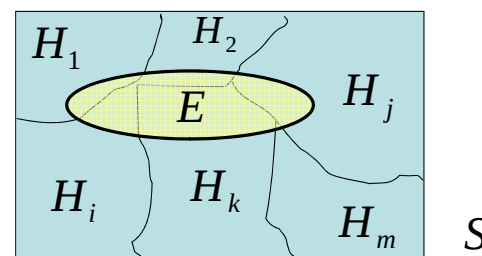
e  $H_i \cap H_j = \emptyset \quad \forall i \neq j$

Incompatibili 2 a 2

Sia  $E$  un evento incluso in  $S$ , cioè

$$E \subset S$$

### Il teorema di Bayes



$$E = E \cap S = E \cap (H_1 \cup H_2 \cup \dots \cup H_m) = \\ = (E \cap H_1) \cup (E \cap H_2) \cup \dots \cup (E \cap H_m)$$



## Il teorema di Bayes

$$\begin{aligned}
 P(E) &= P[(E \cap H_1) \cup (E \cap H_2) \cup \dots \cup (E \cap H_m)] = \\
 &= P(E \cap H_1) + P(E \cap H_2) + \dots + P(E \cap H_m) = \\
 &= \sum_{j=1}^m P(E \cap H_j) = \sum_{j=1}^m P(H_j)P(E | H_j)
 \end{aligned}$$

## Il teorema di Bayes

Se siamo interessati a :

$$\begin{aligned}
 &P(H_i | E) ??? \\
 P(H_i | E) &= \frac{P(H_i \cap E)}{P(E)} =
 \end{aligned}$$

$$= \frac{P(H_i)P(E | H_i)}{\sum_{j=1}^m P(H_j)P(E | H_j)}$$

## Il teorema di Bayes

Cosa rappresentano le singole probabilità?

$$P(H_i) = \text{Probabilità a Priori}$$

$$P(E | H_i) = \text{Probabilità Probative o Verosimiglianze}$$

$$P(H_i | E) = \text{Probabilità a Posteriori}$$

Con questo teorema si risolvono i problemi inversi, cioè:  
Dato un effetto, si vuole risalire alla causa!

## Il teorema di Bayes

Ipotizziamo due possibili cause per l'evento E, cioè\_

$$P(H_1 | E) \quad e \quad P(H_0 | E)$$

$$\text{Se } P(H_1 | E) > P(H_0 | E) \Rightarrow$$

$$\Rightarrow \frac{P(H_1)P(E | H_1)}{P(H_0)P(E | H_0) + P(H_1)P(E | H_1)} >$$

$$> \frac{P(H_0)P(E | H_0)}{P(H_0)P(E | H_0) + P(H_1)P(E | H_1)} \Rightarrow$$

## Il teorema di Bayes

$$\Rightarrow \frac{P(H_1)P(E | H_1)}{\sum_j P(H_j)P(E | H_j)} > \frac{P(H_0)P(E | H_0)}{\sum_j P(H_j)P(E | H_j)} \Rightarrow$$

$$\Rightarrow P(H_1)P(E | H_1) > P(H_0)P(E | H_0) \Rightarrow$$

$$\Rightarrow \frac{P(H_1)P(E | H_1)}{P(H_0)P(E | H_0)} > 1$$

## Il teorema di Bayes

$$\frac{P(H_1)}{P(H_0)} = \text{Rapporto tra Probabilità A priori o } a \text{ priori odds}$$

$$\frac{P(E | H_1)}{P(E | H_0)} = \text{Rapporto tra Verosimiglianze o Fattore di Bayes}$$

## Il teorema di Bayes

$$\text{Se } \frac{P(H_1)}{P(H_0)} \sim 1 \Rightarrow \text{Il fattore di Bayes è } > 1 \text{ ed è quello che ci fa propendere per } H_1$$

$$\text{Se } \frac{P(E | H_1)}{P(E | H_0)} \sim 1 \Rightarrow \text{La scelta per } H_1 \text{ dipende principalmente dall' "a priori odds" che risulta } > 1$$