



UNIVERSITÀ DEGLI STUDI DI NAPOLI  
**FEDERICO II**

Corso di Laurea in Informatica

# **Appunti di Calcolo delle Probabilità e Statistica**

Prof. Aniello Buonocore

A cura di Vincenzo Manno.

# Spazi di probabilità

## 1. Introduzione

**Fenomeni aleatori** (o **casuali**) – fenomeni dei quali non è possibile avere sotto controllo tutti i fattori che ne determinano lo svolgimento, e dei quali, dunque, è impossibile prevedere il risultato.

$(\Omega, A, P)$  – **Spazio di probabilità** usato per descrivere un esperimento aleatorio

- $\Omega$  – **Spazio campione** (contiene tutti i possibili risultati dell'esperimento)
  - $\omega$  – generico evento detto anche **punto campione**
- $A$  – Famiglia degli eventi “interessanti” per l'esperimento ( $A \subseteq \mathcal{P}(\Omega)$ )
- $P$  – funzione che fornisce la probabilità che un evento si verifichi ( $P: A \rightarrow \mathbb{R}^+$ )

Per modellizzare dal punto di vista matematico un fenomeno aleatorio bisogna:

1. Costruire lo spazio campione  $\Omega$
2. Stabilire una famiglia  $A$  di sottoinsiemi di  $\Omega$  che rappresentano la famiglia di eventi “interessanti” per l'esperimento
3. Definire una funzione  $P$  che fornisca le valutazioni di probabilità per i singoli eventi

(1.1) **Assiomi di Kolmogorov** – Una famiglia  $A$  di sottoinsiemi di  $\Omega$  è detta una  **$\sigma$ -algebra degli eventi** se verifica le seguenti proprietà:

- (i)  $\Omega \in A$
- (ii)  $A \in A \Rightarrow A^c \in A$
- (iii) Se  $\{A_n\}_{n \in \mathbb{N}}$  è una successione di elementi di  $A$ , allora  $\bigcup_n A_n \in A$

**Osservazioni sugli assiomi di Kolmogorov**

- (a)  $A \in A$  si può esprimere a parole dicendo semplicemente che “ $A$  è un evento”
- (b)  $\Omega$  è detto anche **evento certo**
- (c) Se  $A$  è un evento,  $A^c$  (che è un evento per l'assioma (ii)) si chiama **evento contrario** di  $A$
- (d) Dagli assiomi (i) e (ii) segue che  $\emptyset = \Omega^c \in A$ . Il sottoinsieme  $\emptyset$  è detto **evento impossibile**
- (e) L'assioma (iii) richiede in particolare che, se una famiglia finita di sottoinsiemi di  $\Omega$  è “interessante” per l'esperimento che stiamo considerando, anche la loro unione lo sia (ciò si esprime dicendo che  $A$  è **stabile rispetto all'unione finita**). Più precisamente, nell'assioma (iii) si vuole che sia un evento anche l'unione di ogni successione numerabile di eventi ( $A$  è **stabile rispetto all'unione numerabile**)
- (f) Dagli assiomi (ii) e (iii) e dalle leggi di De Morgan discende facilmente la seguente proprietà: **stabilità di  $A$  rispetto all'intersezione numerabile**

(iv) Se  $\{A_n\}_n$  è una successione numerabile di elementi di  $A$ , allora  $\bigcap_n A_n \in A$  cioè:

$$\{A_n\}_n \Rightarrow \bigcap_{n=1}^{\infty} A_n \in A$$

Grazie alle formule di De Morgan  $\bigcap_{n=1}^{\infty} A_n$  può essere scritta come  $(\bigcup_{n=1}^{\infty} A_n^c)^c$

Analizzando quest'ultima espressione notiamo che:

- ogni  $A_n^c \in A$  (per l'assioma (ii)<sup>1</sup>)
- l'unione degli  $A_n^c \in A$ , in quanto è l'unione di elementi appartenenti ad  $A$  (assioma (iii)<sup>1</sup>)
- il complemento dell'unione degli  $A_n^c \in A$  (per l'assioma (ii)<sup>1</sup>)

(1.2) Siano  $\Omega$  un insieme,  $A$  una  $\sigma$ -algebra di sottoinsiemi di  $\Omega$ . Un'applicazione  $P: A \rightarrow R^+$  si chiama **probabilità** (su  $A$ ) se verifica le due condizioni seguenti:

- $P(\Omega) = 1$
- Se  $\{A_n\}_{n \in N}$  è una successione di elementi di  $A$  due a due disgiunti (cioè, tali che  $\forall i \neq j$  si abbia  $A_i \cap A_j = \emptyset$ ) risulta  $P(\bigcup_n A_n) = \sum_n P(A_n)$  (proprietà di **finita additività**; nel caso di una successione numerabile, si parla di  **$\sigma$ -additività**)

**Nota:** eventi disgiunti vengono anche detti **eventi incompatibili**.

(1.3) **Teorema** –  $P(\emptyset) = 0$

Dimostrazione – Scriviamo ora l'insieme  $\emptyset$  come unione di un numero infinito di insiemi  $\emptyset$ , cioè  $\emptyset = \emptyset \cup \emptyset \cup \dots \cup \emptyset \dots$

Applicando l'assioma (ii)<sup>2</sup> si ha

$$P(\emptyset) = P\left(\bigcup_{n=1}^{\infty} \emptyset_n\right) = \sum_{n=1}^{\infty} P(\emptyset_n)$$

dove, per  $n = 1, 2, \dots$ ,  $P(\emptyset_1) = P(\emptyset_2) = \dots$

Ponendo  $P(\emptyset_n) = \varepsilon$ , si ha

$$(*) \quad P(\emptyset) = \sum_{n=1}^{\infty} \varepsilon$$

Per  $\varepsilon > 0$ , si ha che la serie (\*) diverge (ciò non può verificarsi, in quanto andrebbe contro l'assioma (i)<sup>2</sup>). Dunque l'unico valore che  $\varepsilon$  può assumere è 0 e si ha che

$$P(\emptyset) = \sum_{n=1}^{\infty} 0 = 0$$

<sup>1</sup> – Assiomi di Kolmogorov per le  $\sigma$ -algebre degli eventi (Assiomi (1.1))

<sup>2</sup> – Assiomi di Kolmogorov per la probabilità (Assiomi (1.2))

(1.4) **Teorema** – La  $\sigma$ –additività comporta anche la finita additività

Dimostrazione – Siano

$$B_1 = A_1$$

$$B_2 = A_2$$

...

$$B_n = A_n$$

$$B_{n+1} = B_{n+2} = \dots = \emptyset$$

sequenze di eventi, dove  $B_i \cap B_j = \emptyset, \forall i \neq j$ .

Per l'assioma (ii)<sup>1</sup> abbiamo

$$(*) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} P(B_k)$$

Scomponendo la serie (\*) nella somma di due serie, in cui la prima va da 1 a  $n$  e la seconda va da  $n+1$  a  $\infty$ , si ha

$$(**) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^n P(B_k) + \sum_{k=n+1}^{\infty} P(B_k)$$

La serie  $\sum_{k=n+1}^{\infty} P(B_k)$  è uguale a 0 in quanto ogni  $B_k$ , con  $k = n+1, n+2, \dots$ , è pari a  $\emptyset$  e la somma di

infiniti  $P(\emptyset)$  è uguale a 0 (vedi (1.3))

Da qui, sostituendo nella (\*) si ha

$$(***) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^n P(B_k)$$

Dalla (\*\*\*) scomponiamo l'unione in

$$(***) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = P\left(\bigcup_{k=1}^n B_k \cup \bigcup_{k=n+1}^{\infty} B_k\right)$$

La  $\bigcup_{k=n+1}^{\infty} B_k$  è uguale a  $\emptyset$  in quanto ogni  $B_k$ , con  $k = n+1, n+2, \dots$ , è pari a  $\emptyset$

Dunque  $P\left(\bigcup_{k=1}^{\infty} B_k\right) = P\left(\bigcup_{k=1}^n B_k\right) = \sum_{k=1}^n P(B_k)$ . Essendo  $B_i = A_i$ , per  $i = 1, 2, \dots, n$ , si ha

$$P\left(\bigcup_{k=1}^n B_k\right) = P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k) = \sum_{k=1}^n P(B_k)$$

<sup>1</sup> – Assioma di Kolmogorov per la probabilità (Assiomi (1.2))

**Nota:** Nel caso di insiemi  $\Omega$  a cardinalità finita o numerabile, tutti i sottoinsiemi di  $\Omega$  sono considerati eventi, cioè si pone sempre  $\mathcal{A} = \mathcal{P}(\Omega)$ .

## 2. Alcune conseguenze degli assiomi

Sia  $(\Omega, \mathcal{A}, P)$  uno spazio di probabilità. Siano  $A$  e  $B$  due eventi. Possiamo scrivere dunque

$$B = B \cup \emptyset = B \cup (A \cap A^c) = (A \cap B) \cup (A^c \cap B)$$

Gli eventi  $(A \cap B)$  e  $(A^c \cap B)$  sono evidentemente disgiunti, pertanto è possibile applicare la proprietà additiva di  $P$ , e abbiamo

$$(2.1) \quad P(B) = P(A \cap B) + P(A^c \cap B)$$

Questa formula è di per sé molto importante: capita spesso di non saper calcolare direttamente la probabilità di un evento (in questo caso  $B$ ), ma di saper “spezzare”  $B$  in due parti disgiunte tramite un evento ausiliario (in questo caso  $A$ ) la cui probabilità è facilmente calcolabile.

La (2.1) ha alcune conseguenze importanti:

- (i) Per  $B = \Omega$  e per ogni  $A \in \mathcal{A}$ , ricordando l’assioma (i) di  $P$ , la (2.1) diventa

$$1 = P(\Omega) = P(A) + P(A^c)$$

ovvero

$$(2.2) \quad P(A^c) = 1 - P(A)$$

La (2.2) permette di calcolare la probabilità dell’evento contrario di  $A$  se è nota la probabilità di  $A$ .

- (ii) Per  $A \subseteq B$ , la (2.1) diventa

$$(2.3) \quad P(B) = P(A) + P(A^c \cap B)$$

poiché  $P(A^c \cap B) \geq 0$ , si ricava per  $A \subseteq B$

$$(2.4) \quad P(A) \leq P(B)$$

Questa proprietà si chiama proprietà di **isotonia** della probabilità.

Una sua semplice conseguenza è il fatto che ogni evento ha probabilità inferiore o uguale a 1: infatti, se  $A$  è un evento, dalla ovvia relazione  $A \subseteq \Omega$  segue

$$P(A) \leq P(\Omega) = 1$$

In altre parole, ogni funzione di probabilità assume valori nell’intervallo  $[0, 1]$ .

- (iii) Se  $A \subseteq B$ , si usa scrivere

$$A^c \cap B = B \setminus A$$

e la (2.3) diventa

$$P(B \setminus A) = P(B) - P(A)$$

- (iv) Siano ora A e B due eventi qualsiasi. Possiamo scrivere

$$A \cup B = A \cup (A^c \cap B)$$

essendo i due eventi A e  $(A^c \cap B)$  disgiunti, dalla proprietà di additività della probabilità e dalla (2.1) si deduce

$$(2.5) \quad P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B)$$

Questa formula permette di calcolare la probabilità dell'unione di due eventi A e B, a patto di conoscere già le probabilità degli eventi A, B e della loro intersezione.

La (2.5) è un caso particolare di una formula (nota come la **formula di inclusione-esclusione**) che si utilizza per il calcolo della probabilità dell'unione di una famiglia finita di eventi. Per completezza diamo anche la formula per il caso di tre eventi A, B e C

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

per la (2.5) si ha

$$P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C)$$

applicando la (2.5) a  $P(A \cup B)$

$$P((A \cup B) \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P((A \cup B) \cap C)$$

applicando la proprietà distributiva dell'intersezione rispetto all'unione a  $(A \cup B) \cap C$

$$P((A \cup B) \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P((A \cap C) \cup (B \cap C))$$

applicando la (2.5) a  $P((A \cap C) \cup (B \cap C))$

$$P((A \cup B) \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - [P((A \cap C) \cup (B \cap C)) - P(A \cap B \cap C)]$$

ovvero

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P((A \cap C) \cup (B \cap C)) + P(A \cap B \cap C)$$

- (v) Sia  $\{A_n\}_n$  una successione di eventi. Poiché

$$\cup_n A_n = (\cap_n A_n^c)^c$$

dalla (2.2) si ricava

$$(2.6) \quad P(\cup_n A_n) = P((\cap_n A_n^c)^c) = 1 - P(\cap_n A_n^c)$$

Questa formula è spesso utile perchè tipicamente la probabilità di un'intersezione di eventi risulta più facile da calcolare della probabilità di un'unione.

Tutte le proprietà viste finora sono conseguenza della sola proprietà di additività finita della probabilità. Queste ultime due, invece, utilizzano la  $\sigma$ -additività.

- (vi) (***proprietà di passaggio al limite sulle successioni crescenti***) Sia  $\{A_n\}_n$  una successione crescente di eventi (cioè,  $\forall n \in N$ , risulta  $A_n \subseteq A_{n+1}$ ). Si ha allora

$$P(\cup_n A_n) = P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

- (vii) (***proprietà di passaggio al limite sulle successioni decrescenti***) Sia  $\{A_n\}_n$  una successione decrescente di eventi (cioè,  $\forall n \in N$ , risulta  $A_n \supseteq A_{n+1}$ ). Si ha allora

$$P(\cap_n A_n) = P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

### 3. Spazi di probabilità uniformi

Supponiamo di essere in presenza di un esperimento aleatorio che si svolge in **condizioni di equiprobabilità** (cioè, come nel caso dei lanci della moneta equilibrata, in modo che non ci sia motivo di pensare che un particolare risultato sia più o meno frequente degli altri), e sia  $\Omega$  lo spazio campione associato ad esso. La condizione di equiprobabilità impone evidentemente che ci sia

$$P(\omega) = \text{costante} \equiv p, \quad \forall \omega \in \Omega$$

Determiniamo allora la costante  $p$ . Si ha l'eguaglianza ovvia

$$\Omega = \cup_{\omega \in \Omega} \{\omega\}$$

per gli assiomi (i)<sup>1</sup> e (ii)<sup>1</sup> si deduce

$$1 = P(\Omega) = \sum_{\omega \in \Omega} P(\omega) = \sum_{\omega \in \Omega} p = p (\text{card } \Omega)$$

da questa relazione segue allora

$$p = \frac{1}{\text{card } \Omega}$$

Questo determina la probabilità di ciascun punto campione. Per trovare ora la probabilità di un qualunque evento  $A$  si ragiona in modo sostanzialmente identico: dall'eguaglianza

$$A = \cup_{\omega \in A} \{\omega\}$$

si deduce

$$(3.1) \quad P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} \frac{1}{\text{card } \Omega} = \frac{\text{card } A}{\text{card } \Omega}$$

Questa formula definisce allora la funzione probabilità su tutta la  $\sigma$ -algebra degli eventi.

Sia  $(\Omega, A, P)$  uno spazio di probabilità, tale che  $\text{card } \Omega < +\infty$ .  $(\Omega, A, P)$  si dice uno **spazio di probabilità uniforme** se

$$P(\omega) = \frac{1}{\text{card } \Omega}$$

per ogni  $\omega \in \Omega$ .

La formula (3.1) dice semplicemente che, nel caso di uno spazio uniforme, la probabilità di un evento si calcola come il rapporto tra il “numero dei casi favorevoli” e il “numero dei casi possibili”. E' bene ricordare, comunque, che solo in alcuni casi lo spazio di probabilità da costruire sarà uniforme.

<sup>1</sup> – Assioma di Kolmogorov per la probabilità (Assiomi (1.2))



#### 4. Probabilità condizionale e indipendenza

(4.1) Siano  $A, B$  due eventi, con  $P(B) > 0$ . Si chiama **probabilità condizionale di  $A$ , dato  $B$**  (oppure “*sapendo che si è verificato  $B$* ”) la quantità

$$(4.2) \quad P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Si sottolinea il fatto che la scrittura  $P(A | B)$  non rappresenta la probabilità dell'evento “ $A | B$ ”, ma solo un modo per indicare il rapporto che compare al secondo membro della (4.2).

(4.3) **Teorema** – Sia  $B$  un evento, con  $P(B) > 0$ . L'applicazione  $Q: A \rightarrow R^+$  definita da  $Q(A) = P(A | B)$  è una probabilità.

Dimostrazione – Dalla definizione di probabilità (1.2), l'applicazione  $Q$  è una probabilità (su  $A$ ) se verifica le seguenti due condizioni:

- (i)  $Q(\Omega) = 1$
- (ii) Se  $\{A_n\}_n$  è una successione di elementi di  $A$  due a due disgiunti (cioè, tali che  $\forall i \neq j$  si abbia  $A_i \cap A_j = \emptyset$ ) risulta  $Q(\cup_n A_n) = \sum_n Q(A_n)$
- (i) per come è stata definita l'applicazione

$$Q(\Omega) = P(\Omega | B)$$

per la (4.2) si ha

$$P(\Omega | B) = \frac{P(\Omega \cap B)}{P(B)}$$

essendo  $\Omega \cap B = B$

$$\frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

- (ii) per come è stata definita l'applicazione

$$Q(\cup_n A_n) = P(\cup_n A_n | B)$$

cioè

$$P(\cup_n A_n | B) = P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) | B)$$

applicando la (4.2) otteniamo

$$P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) | B) = \frac{P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) \cap B)}{P(B)}$$

per la proprietà distributiva di  $\cap$  rispetto a  $\cup$ , abbiamo

$$\frac{P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B) \cup \dots)}{P(B)}$$

gli eventi  $A_i \cap B$ , per  $i = 1, 2, \dots$ , sono evidentemente disgiunti; dunque è possibile applicare l'assioma (ii)<sup>1</sup> e si ha

$$\frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B) \cup \dots)}{P(B)} = \frac{P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B) + \dots}{P(B)}$$

le varie  $P(A_i \cap B)$  possono essere raggruppate tutte in una serie, ovvero

$$(*) \quad \frac{P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B) + \dots}{P(B)} = \sum_n \frac{P(A_n \cap B)}{P(B)}$$

per la (4.2) si ha

$$(**) \quad \frac{P(A_n \cap B)}{P(B)} = P(A_n | B)$$

dove, per l'ipotesi del teorema,

$$(***) \quad P(A_n | B) = Q(A_n)$$

applicando, infine, la (\*\*) e (\*\*\*) alla (\*) si ottiene

$$\sum_n \frac{P(A_n \cap B)}{P(B)} = \sum_n P(A_n | B) = \sum_n Q(A_n)$$

Può capitare che l'informazione supplementare "si è verificato B" non cambi la nostra valutazione della probabilità del verificarsi di A.

Siano A e B due eventi, con  $P(B) > 0$ . A e B si dicono *indipendenti (tra loro)* se

$$(4.4) \quad P(A | B) = P(A)$$

Usando la (4.2), la (4.4) si può scrivere nella forma

$$\frac{P(A \cap B)}{P(B)} = P(A)$$

o anche, moltiplicando per  $P(B)$ ,

$$(4.5) \quad P(A \cap B) = P(A)P(B)$$

La (4.5) è chiaramente simmetrica in A e B (del resto lo è anche l'espressione A e B sono indipendenti (tra loro) usata nella definizione, mentre non è affatto evidente che lo sia anche la (4.4)). Inoltre, mentre la (4.4) perde di significato se B è un evento con probabilità nulla, la (4.5) ha ancora senso. Essa risulta più comoda come definizione di indipendenza; dunque sostituiremo la definizione provvisoria con la seguente:

Due eventi A e B si dicono *indipendenti (tra loro)* se vale la (4.5)

Sarà necessario estendere la definizione di indipendenza ad una famiglia qualsiasi di eventi. Per iniziare, consideriamo il caso di tre eventi A, B e C. Per analogia con il caso di due eventi, potremmo pensare che sia corretto dire che essi sono indipendenti se

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

Purtroppo questa condizione non garantisce che gli eventi siano anche “due a due” indipendenti, come ci dice invece l’intuizione. La definizione ragionevole è allora la seguente

Tre eventi A, B e C si dicono **indipendenti** se valgono tutte le seguenti condizioni

- (i)  $P(A \cap B \cap C) = P(A)P(B)P(C)$
- (ii)  $P(A \cap B) = P(A)P(B)$
- (iii)  $P(A \cap C) = P(A)P(C)$
- (iv)  $P(B \cap C) = P(B)P(C)$

La definizione precedente si generalizza in modo ovvio ad una famiglia di qualsiasi eventi: essi si diranno indipendenti se per ogni sottofamiglia finita la probabilità dell’intersezione è uguale al prodotto delle singole probabilità; in termini precisi

(4.6) Gli eventi  $\{A_i\}_{i \in I}$  si dicono **indipendenti (tra loro)** se per ogni  $k \in N$  e per ogni successione finita  $i_1, i_2, \dots, i_k$  di indici di  $I$ , si ha

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

**Nota:** L’ipotesi di indipendenza consente di costruire un particolare tipo di spazio di probabilità, utilizzabile nella pratica in una varietà di situazioni solo apparentemente differenti tra loro.

(4.7) Siano A e B due eventi. Le affermazioni seguenti sono equivalenti:

- (a) A e B sono indipendenti
- (b) A e  $B^c$  sono indipendenti
- (c)  $A^c$  e B sono indipendenti
- (d)  $A^c$  e  $B^c$  sono indipendenti

Dimostrazione – Supponiamo che i due eventi siano indipendenti (affermazione (a)), ovvero che

$$(*) \quad P(A \cap B) = P(A)P(B)$$

Dimostriamo che le rimanenti affermazioni sono equivalenti.

(b) Scomponiamo  $P(A)$ , come in (2.1), tramite l’evento ausiliario B ed abbiamo

$$(b.1) \quad P(A) = P(A \cap B) + P(A \cap B^c)$$

sostituendo la (\*) in (b.1) si ha

$$P(A \cap B^c) = P(A) - P(A)P(B)$$

mettendo in evidenza  $P(A)$  al secondo membro, otteniamo

$$(b.2) \quad P(A \cap B^c) = P(A)(1 - P(B))$$

per la (2.2), la (b.2) diventa

$$P(A \cap B^c) = P(A)P(B^c)$$

(c) Scomponiamo  $P(B)$ , come in (2.1), tramite l'evento ausiliario  $A$  ed abbiamo

$$(c.1) \quad P(B) = P(A \cap B) + P(A^c \cap B)$$

sostituendo la (\*) in (c.1) si ha

$$P(A^c \cap B) = P(B) - P(A)P(B)$$

mettendo in evidenza  $P(B)$  al secondo membro, otteniamo

$$(c.2) \quad P(A^c \cap B) = P(B)(1 - P(A))$$

per la (2.2), la (c.2) diventa

$$P(A^c \cap B) = P(A^c)P(B)$$

(d) Dalle formule di De Morgan otteniamo

$$P(A^c \cap B^c) = P((A \cup B)^c)$$

applicando la (2.2)

$$P(A^c \cap B^c) = 1 - P(A \cup B)$$

dalla (2.5) e (\*) abbiamo

$$P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A)P(B)$$

applicando la (2.2) a  $(1 - P(A))$

$$P(A^c \cap B^c) = P(A^c) - P(B) + P(A)P(B)$$

mettendo in evidenza  $-P(B)$  si ha

$$P(A^c \cap B^c) = P(A^c) - P(B)(1 - P(A))$$

applicando la (2.2) a  $(1 - P(A))$

$$P(A^c \cap B^c) = P(A^c) - P(B)P(A^c)$$

mettendo in evidenza  $P(A^c)$

$$P(A^c \cap B^c) = P(A^c)(1 - P(B))$$

applicando la (2.2) a  $(1 - P(B))$

$$P(A^c \cap B^c) = P(A^c)P(B^c)$$

## 5. Schema delle prove indipendenti

Supponiamo di dover eseguire  $n$  volte (in condizioni di indipendenza) un certo esperimento che produce due soli possibili risultati, che chiameremo convenzionalmente **successo** e **insuccesso** e che indicheremo rispettivamente con i simboli 1 e 0. Supponiamo inoltre di sapere che il successo (risp. l'insuccesso) si presenta con probabilità  $p$  (risp.  $1 - p$ ) nella generica ripetizione dell'esperimento (con  $0 < p < 1$ ). Lo spazio campione è il seguente

$$\Omega = \{0, 1\}^n$$

Sia dunque  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$  (con  $\omega_i \in \{0, 1\}$ ) un elemento di  $\Omega$ ; osserviamo che le quantità

$$\sum_{i=1}^n \omega_i, \quad n - \sum_{i=1}^n \omega_i$$

Non sono altro che il numero di simboli uguali a 1 e il numero i simboli uguali a 0 che formano la sequenza  $\omega$ . Pertanto avremo

$$P(\omega) = p^{(\sum_{i=1}^n \omega_i)} (1 - p)^{(n - \sum_{i=1}^n \omega_i)}$$

## 6. La formula di Bayes

(6.1) Siano  $(\Omega, A, P)$  uno spazio di probabilità fissato, e  $A_1, \dots, A_n$   $n$  eventi in  $A$ . Si dice che  $A_1, \dots, A_n$  è una **partizione** di  $(\Omega, A, P)$  se valgono entrambe le seguenti proprietà:

- (i)  $\bigcup_{i=1}^n A_i = \Omega$
- (ii)  $A_i \cap A_j = \emptyset \quad \forall (i, j) \text{ con } i \neq j$

Gli eventi di una partizione vanno pensati come una successione di  $n$  eventualità che possono presentarsi durante lo svolgimento dell'esperimento; esse esauriscono tutte le possibilità (sicuramente una di esse si verifica) e si escludono a vicenda (due di esse non possono verificarsi contemporaneamente).

(6.2) **Teorema (Formula di Bayes)** – Sia  $(\Omega, A, P)$  uno spazio di probabilità. Sia  $A_1, \dots, A_n$  una partizione di  $\Omega$ , tale che  $P(A_i) > 0$  per ogni  $i = 1, 2, \dots, n$ . Sia infine  $B$  un evento con  $P(B) > 0$ . Allora per ogni  $k = 1, 2, \dots, n$  risulta

$$P(A_k | B) = \frac{P(B | A_k)P(A_k)}{\sum_{i=1}^n P(B | A_i)P(A_i)}$$

Dimostrazione – Per la (4.2) abbiamo che

$$(*) \quad P(A_k | B) = \frac{P(A_k \cap B)}{P(B)}$$

Ora soffermiamoci sul numeratore e sul denominatore della (\*)

- (numeratore) per la proprietà simmetrica dell'intersezione

$$P(A_k \cap B) = P(B \cap A_k)$$

applicando la (4.2) a  $P(B \cap A_k)$  si ottiene

$$(**) \quad P(B \cap A_k) = P(B | A_k)P(A_k)$$

- (denominatore) possiamo scrivere

$$B = B \cap \Omega$$

per il punto (i) della definizione (6.1) otteniamo che

$$B = B \cap \Omega = B \cap \left( \bigcup_{i=1}^n A_i \right)$$

applicando la proprietà distributiva di  $\cap$  rispetto a  $\cup$ , abbiamo

$$B \cap \left( \bigcup_{i=1}^n A_i \right) = \left( \bigcup_{i=1}^n (B \cap A_i) \right)$$

poiché gli eventi  $(B \cap A_i)$  sono tra loro a due a due disgiunti, per l'assioma di additività della probabilità si ottiene

$$(***) \quad P(B) = \sum_{i=1}^n P(B \cap A_i)$$

come abbiamo fatto in precedenza per  $P(B \cap A_k)$  (vedi (\*\*)) si può scrivere

$$P(B \cap A_i) = P(B | A_i)P(A_i)$$

e quindi la (\*\*\*) diventa

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i)$$

applicando la (\*\*) e la (\*\*\*\*) alla (\*) si ricava allora

$$P(A_k | B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B | A_k)P(A_k)}{\sum_{i=1}^n P(B | A_i)P(A_i)}$$

ovvero la formula di Bayes che cercavamo, in cui

1.  $P(A_k | B)$  è detta ***probabilità a posteriori***
2.  $P(B | A_k)$  è detta ***probabilità di verosimiglianza***
3.  $P(A_k)$  è detta ***probabilità a priori***

## 1 Esempio 4.1.1 del libro di testo consigliato

L'esempio è esemplificativo della descrizione formale di un *esperimento aleatorio* e dei *numeri non deterministici*.

Si consideri l'esperimento  $\mathcal{E}_2$  che consiste nel lanciare due dadi *onesti* (ossia, non sbilanciati) di colore differente in maniera tale che si può parlare di primo e secondo dado. Lo *spazio campione*, ovvero l'insieme dei possibili risultati di  $\mathcal{E}_2$ , è:

$$\Omega_2 := \{(i, j) : i, j = 1, 2, \dots, 6\}.$$

Ne risulta che  $|\Omega_2| = 6^2$ , ovvero che in  $\Omega_2$  ci sono 36 coppie ordinate ognuna delle quali è un *punto campione*. D'altra parte, l'ipotesi dell'onestà dei due dadi porta a ritenere che

$$i, j = 1, 2, \dots, 6, \quad \mathbb{P}_2(\{(i, j)\}) = \frac{1}{6^2} = \frac{1}{36},$$

ovvero che in questo specifico contesto è possibile utilizzare la definizione di probabilità introdotta da Laplace (rapporto tra gli esiti favorevoli e gli esiti possibili). Allora, considerando *evento* ogni sottoinsieme di  $\Omega_2$ , si ha:

$$A \in \mathcal{P}(\Omega_2), \quad \mathbb{P}_2(A) = \frac{|A|}{36}. \quad (1)$$

In definitiva, la terna  $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_2)$  con  $\mathbb{P}_2$  definita nella (1) descrive completamente e in maniera formale l'esperimento  $\mathcal{E}_2$ .

Raramente, però, si è interessati al risultato (esito, punto campione) dell'esperimento ma, piuttosto, il valore rilevante è una sua funzione. In questo contesto, ad esempio, ha senso considerare la somma dei punteggi dei due dadi:

$$Z_2 : (i, j) \in \Omega_2 \rightarrow (i + j) \in \mathbb{R}.$$

La funzione  $Z_2$  assume con probabilità positiva i valori:  $2, 3, \dots, 11, 12$ . Per ognuno di tali valori è possibile determinare agevolmente la sua *controimmagine* tramite la funzione  $Z_2$ :

$$\begin{aligned} \{Z_2 = 2\} &= \{(1, 1)\}, & \{Z_2 = 12\} &= \{(6, 6)\}, \\ \{Z_2 = 3\} &= \{(1, 2), (2, 1)\}, & \{Z_2 = 11\} &= \{(5, 6), (6, 5)\}, \\ \{Z_2 = 4\} &= \{(1, 3), (2, 2), (3, 1)\}, & \{Z_2 = 10\} &= \{(4, 6), (5, 5), (6, 4)\}, \\ \{Z_2 = 5\} &= \{(1, 4), (2, 3), (3, 2), (4, 1)\}, & \{Z_2 = 9\} &= \{(3, 6), (4, 5), (5, 4), (6, 3)\}, \\ \{Z_2 = 6\} &= \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}, & \{Z_2 = 8\} &= \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}, \\ \{Z_2 = 7\} &= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}. \end{aligned} \quad (2)$$

Si osservi che la classe costituita dai sottoinsiemi individuati nella (2) costituisce una *partizione* di  $\Omega_2$ .

Dalle (1) e (2) se ne deduce che

$$\begin{aligned} \mathbb{P}_2(Z_2 = 2) &= \mathbb{P}_2(Z_2 = 12) = \frac{1}{36}, & \mathbb{P}_2(Z_2 = 3) &= \mathbb{P}_2(Z_2 = 11) = \frac{2}{36}, \\ \mathbb{P}_2(Z_2 = 4) &= \mathbb{P}_2(Z_2 = 10) = \frac{3}{36}, & \mathbb{P}_2(Z_2 = 5) &= \mathbb{P}_2(Z_2 = 9) = \frac{4}{36}, \\ \mathbb{P}_2(Z_2 = 6) &= \mathbb{P}_2(Z_2 = 8) = \frac{5}{36}, & \mathbb{P}_2(Z_2 = 7) &= \frac{6}{36}, \end{aligned} \quad (3)$$

e, inoltre, risulta

$$\sum_{s=2}^{12} \mathbb{P}_2(Z_2 = s) = \frac{2(1 + 2 + 3 + 4 + 5) + 6}{36} = 1.$$



## 2 Estensione al caso di tre dadi

Si consideri l'esperimento  $\mathcal{E}_3$  che consiste nel lanciare tre dadi onesti di colore differente in maniera tale che si possa parlare di primo, secondo e terzo dado. Lo spazio campione, ovvero l'insieme dei possibili risultati di  $\mathcal{E}_3$ , è:

$$\Omega_3 := \{(i, j, k) : i, j, k = 1, 2, \dots, 6\}.$$

Ne risulta che  $|\Omega_3| = 6^3$ , ovvero che in  $\Omega_3$  ci sono 216 coppie ordinate ognuna delle quali è un punto campione. D'altra parte, l'ipotesi dell'onestà dei tre dadi porta a ritenere che

$$i, j, k = 1, 2, \dots, 6, \quad \mathbb{P}_3(\{(i, j, k)\}) = \frac{1}{6^3} = \frac{1}{216},$$

ovvero che è possibile utilizzare la definizione di probabilità introdotta da Laplace. Allora, considerando evento ogni sottoinsieme di  $\Omega_3$ , si ha:

$$A \in \mathcal{P}(\Omega_3), \quad \mathbb{P}_3(A) = \frac{|A|}{216}. \quad (4)$$

In definitiva, la terna  $(\Omega_3, \mathcal{P}(\Omega_3), \mathbb{P}_3)$  con  $\mathbb{P}_3$  definita nella (4) descrive completamente e in maniera formale l'esperimento  $\mathcal{E}_3$ .

In questo contesto, ad esempio, ha senso considerare la somma dei punteggi dei tre dadi:

$$Z_3 : (i, j, k) \in \Omega_3 \rightarrow (i + j + k) \in \mathbb{R}.$$

La funzione  $Z_3$  assume con probabilità positiva i seguenti valori: 3, 4, ..., 17, 18. Per ognuno di tali valori è possibile determinare agevolmente la sua controimmagine tramite la funzione  $Z_3$ . A tale scopo, per sintetizzare le formule, è utile individuare le classi di terne compatibili con la somma in questione e che sono diverse per almeno un punteggio con le rispettive dimensioni. Ad esempio, per  $\{Z_3 = 9\}$  ci sono le seguenti 6 classi terne:

$$[(1, 2, 6); \textcolor{red}{6}], [(1, 3, 5); \textcolor{red}{6}], [(1, 4, 4); \textcolor{red}{3}], [(2, 2, 5); \textcolor{red}{3}], [(2, 3, 4); \textcolor{red}{6}], [(3, 3, 3); \textcolor{red}{1}].^1$$

Dopo di ciò, si ha:

$$\mathbb{P}_3(Z_3 = 9) = \frac{\textcolor{red}{6} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{3} + \textcolor{red}{6} + \textcolor{red}{1}}{6^3} = \frac{25}{216}, \quad (5)$$

Allo stesso modo, in  $\{Z_3 = 10\}$  ci sono le seguenti 6 classi terne:

$$[(1, 3, 6); \textcolor{red}{6}], [(1, 4, 5); \textcolor{red}{6}], [(2, 2, 6); \textcolor{red}{3}], [(2, 3, 5); \textcolor{red}{6}], [(2, 4, 4); \textcolor{red}{3}], [(3, 3, 4); \textcolor{red}{3}],$$

e, pertanto, risulta:

$$\mathbb{P}_3(Z_3 = 10) = \frac{\textcolor{red}{6} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{3}}{6^3} = \frac{27}{216}. \quad (6)$$

Le formule (5) e (6) forniscono la soluzione al problema posto dai gentiluomini fiorentini a Galileo Galilei: in totale, ci sono 25 terne di punteggi che fanno realizzare il 9 e 27 terne di punteggi che fanno realizzare il 10.

Si lascia allo studente la verifica che

$$\sum_{s=3}^{18} \mathbb{P}_3(Z_3 = s) = 1.$$

<sup>1</sup>La scrittura  $[(1, 2, 6); \textcolor{red}{6}]$  indica le **sei** terne che si ottengono permutando i punteggi 1, 2 e 6 in tutti i modi possibili; analogamente, la scrittura  $[(1, 4, 4); \textcolor{red}{3}]$  indica le **tre** terne che si ottengono permutando i punteggi 1, 4 e 4 in tutti i modi possibili.

*Osservazione 1.* Il numero aleatorio  $Z_2$  può essere considerato anche nello spazio di probabilità  $(\Omega_3, \mathcal{P}(\Omega_3), \mathbb{P}_3)$ :

$$Z_2 : (i, j, k) \in \Omega_3 \rightarrow (i + j + 0 \cdot k) \in \mathbb{R}.$$

Ovviamente, questo comporta un aggravio nelle formule in quanto, ad esempio,

$$\{Z_2 = 2\} = \{(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 1, 4), (1, 1, 5), (1, 1, 6)\} \implies |Z_2 = 2| = 6.$$

Questo fatto non altera i risultati individuati nel paragrafo 1; infatti, relativamente all'evento  $\{Z_2 = 2\}$ , tenendo presente la prima delle formule (3) si ha:

$$\mathbb{P}_3(Z_2 = 2) = \frac{6}{216} = \frac{1}{36} = \mathbb{P}_2(Z_2 = 2).$$

## Richiami di calcolo combinatorio

(1) Dati due insiemi finiti A e B, con  $\text{card } A = m$ ,  $\text{card } B = n$ , si ha

$$\text{card}(A \times B) = m \cdot n$$

(2) Siano  $A_1, \dots, A_n$   $n$  insiemi finiti, con  $\text{card } A_i = m_i$ . Si deduce facilmente per induzione da (1) che

$$\text{card}(A_1 \times A_2 \times \dots \times A_n) = m_1 \cdot m_2 \cdot \dots \cdot m_n$$

(3) E' importante il caso particolare di (2) in cui sia

$$A_1 = A_2 = \dots = A_n = A$$

Posto  $m = \text{card } A$ , si trova

$$\text{card}(A^n) = (\text{card } A)^n = m^n$$

(4) Siano dati due insiemi X e A, con  $\text{card } X = n$ ,  $\text{card } A = m$ . Allora il numero delle applicazioni di X in A è dato da  $m^n$ .

(5) Ogni elemento del prodotto cartesiano  $A^n = A \times A \times \dots \times A$  ( $n$  volte può essere visto come una  $n$ -upla di elementi di A, dei quali, eventualmente, qualcuno sia ripetuto. Le  $n$ -uple di elementi di A si chiamano anche **sequenze con ripetizione** (di elementi di A) di lunghezza  $n$ . Il numero di tali sequenze è dunque, come abbiamo appena visto,  $(\text{card } A)^n$ .

(6) Dato un insieme A formato da  $m$  oggetti, siamo interessati anche al numero di **sequenze senza ripetizione** (di elementi di A) di lunghezza  $n$  ( $0 < n \leq m$ ). Esse vengono chiamate anche **disposizioni** di  $m$  oggetti a  $n$  a  $n$ . E' facile vedere che questo numero è dato da

$$(m)_n = m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot (m-n+1) = \frac{m!}{(m-n)!}$$

(7) Siano X e A due insiemi come in (4), e supponiamo  $0 < n \leq m$ . Il numero delle applicazioni iniettive di X in A è dato da  $(m)_n$ .

(8) E' importante il caso particolare di (6) in cui sia  $n = m$ ;  $(m)_m$  è il numero di modi diversi in cui si possono disporre gli  $m$  elementi di A, o, come si dice, il numero delle **permutazioni** degli  $m$  oggetti che formano l'insieme A.  $(m)_m$  si indica più frequentemente con il simbolo  $m!$ , e vale

$$m! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot m$$

(9) Caso particolare di (8) si ha quando le permutazioni sono **circolari**. Il numero di quest'ultime è  $m-1!$

(10) Sia  $A$  come in (6). Due sequenze senza ripetizione di elementi di  $A$  possono differire per gli oggetti che le compongono, oppure anche soltanto per l'ordine in cui questi oggetti vengono presi; in molte situazioni, tuttavia, basterà contare il numero delle sequenze che differiscono per la natura degli elementi (identificando cioè quelle formate dagli stessi oggetti, presi in ordine diverso); ci interessa dunque trovare il numero di gruppi costituiti da  $n$  elementi, che si possono formare a partire da un insieme di cardinalità  $m$ . Questi gruppi si chiamano **combinazioni** di  $m$  oggetti  $n$  a  $n$ . Il numero che cerchiamo vale

$$\frac{(m)_n}{n!} = \frac{m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot (m-n+1)}{n!} = \frac{m!}{n!(m-n)!} = \binom{m}{n}$$

$\binom{m}{n}$  si chiama **coefficiente binomiale**  $m, n$ . Osserviamo che

$$\binom{m}{n} = \binom{m}{m-n}$$

(11) Caso particolare di (10) si ha quando le combinazioni sono con ripetizione. Il numero di quest'ultime è dato da

$$\binom{m+n-1}{n} = \frac{(m+n-1)!}{n!(m-1)!}$$

(12) E' necessario dare un senso anche al simbolo  $0!$  (il caso  $m = n = 0$ ) non rientra nella situazione considerata in (8). Possiamo ragionare così: se in (10) prendiamo  $n = m$ , il coefficiente binomiale risultante dovrà essere il numero di gruppi costituiti da  $m$  oggetti, formati a partire da un insieme di cardinalità  $m$ . E' chiaro che di questi gruppi ne potremo formare solo uno (l'insieme dato) e quindi dovremo avere

$$\binom{m}{m} = 1$$

d'altra parte, sviluppando formalmente la formula che definisce il coefficiente binomiale, si trova

$$\binom{m}{m} = \frac{m!}{m!0!} = \frac{1}{0!}$$

Dovremo pertanto porre  $0! = 1$ .

(13) Sia  $A$  come in (6), e siano  $n_1, n_2, \dots, n_r$  degli interi maggiori o uguali a 0, tali che  $n_1 + n_2 + \dots + n_r = m$ . In quanti modi si possono formare  $r$  gruppi (a partire dagli elementi di  $A$ )  $G_1, G_2, \dots, G_r$ , in modo che il gruppo  $G_i$  contenga  $n_i$  elementi?

Per rispondere alla domanda, possiamo pensare di procedere così: formiamo prima un gruppo di  $n_1$  elementi, poi, con gli  $m - n_1$  elementi restanti, un gruppo di  $n_2$ , e così via. Il numero di gruppi possibili è quindi dato da

$$\binom{m}{n_1} \cdot \binom{m-n_1}{n_2} \cdot \dots \cdot \binom{m-n_1-\dots-n_{r-1}}{n_r}$$

Si vede facilmente che la quantità scritta sopra è uguale a

$$\frac{m!}{n_1! \cdot n_2! \cdot \dots \cdot n_r!}$$

e si indica col simbolo

$$\binom{m}{n_1, n_2, \dots, n_r},$$

detto **coefficiente multinomiale**  $m, n_1, n_2, \dots, n_r$ .

Per  $r = 2$  si ha

$$\binom{m}{n_1, n_2} = \binom{m}{n_1} = \binom{m}{n_2}$$

(14) Se  $1 \leq n \leq m-1$  si ha

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n}$$

(15) **Il binomio di Newton.** Siano  $a, b$  due numeri reali,  $n$  un intero non negativo. Si ha allora

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

(16) Un caso particolare della formula precedente: prendendo  $a = b = 1$ , si trova

$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

Ricordando il significato combinatorio dei coefficienti binomiali  $\binom{n}{k}$ , la formula scritta sopra si può interpretare dicendo che il numero dei sottoinsiemi che si possono formare a partire da un insieme di  $n$  elementi è pari a  $2^n$ .

### Schema riassuntivo

Riferimento	Descrizione	Proprietà	Numero
(5)	Sequenze con ripetizioni	$O - R$	$m^n$
(6)	Sequenze senza ripetizioni (o anche disposizioni)	$O - \overline{R}$	$(m)_n$
(8)	Permutazioni (caso particolare di (6))	$O - \overline{R}$	$m!$
(9)	Permutazioni circolari (caso particolare di (8))	$O - \overline{R}$	$m-1!$
(10)	Combinazioni	$\overline{O} - \overline{R}$	$\binom{m}{n}$
(11)	Combinazioni con ripetizione	$\overline{O} - R$	$\binom{m+n-1}{n}$
(13)	Permutazioni con ripetizione	$\overline{O} - R$	$\binom{m}{n_1, n_2, \dots, n_r}$

O – Ordine (cioè due sequenze composte dagli stessi simboli, ma posti in un ordine diverso, non sono considerate la medesima)

R – Ripetizione (è possibile che uno o più elementi vengano ripetuti una o più volte)

# Variabili aleatorie

## 1. Definizioni e prime proprietà

Molto spesso, l'esperimento che viene effettuato produce una quantità numerica  $X$  che deve essere studiata; anzi, si può dire che in moltissime situazioni è proprio la quantità  $X$  l'oggetto interessante per chi esegue l'esperimento, (più che l'esperimento in sé per sé).

Esempio – Un programma informatico produce sequenze aleatorie di lunghezza 5, composte dai simboli 0 e 1. Questo esperimento si modella con uno schema  $n = 5$  prove indipendenti ( $p$  = probabilità di ottenere un simbolo “1”).

E' noto che il tempo impiegato per produrre un simbolo “1” (risp. “0”) è 2 (risp. 3)  $\mu\text{sec}$ . Il tecnico che utilizza il programma è interessato a sapere quanto tempo è necessario per produrre una sequenza; in altre parole, le domande a cui vuole saper rispondere sono del tipo “qual è la probabilità che il programma impieghi più di  $x \mu\text{sec}$ ”, oppure “qual è la probabilità che il programma impieghi un tempo compreso fra  $x \mu\text{sec}$  e  $y \mu\text{sec}$ ”, e così via, (piuttosto che domande come “qual è la probabilità che il programma produca (per esempio) la sequenza (1, 0, 0, 1, 1)”).

Cioè in questa situazione siamo interessati a ottenere informazioni “di tipo probabilistico” sulla quantità  $X$  = tempo (in  $\mu\text{sec}$ ) impiegato dal programma per produrre una sequenza.

Cerchiamo ora di formalizzare domande del tipo “qual è la probabilità che il programma impieghi un tempo compreso fra  $x \mu\text{sec}$  e  $y \mu\text{sec}$ ”. Questa domanda equivale a cercare la probabilità del sottoinsieme costituito dai possibili risultati  $\omega$  dell'esperimento per i quali risulti  $x < X(\omega) \leq y$ , ovvero la probabilità di

$$\{\omega \in \Omega : x < X(\omega) \leq y\}$$

Pertanto una richiesta ragionevole sarà che i sottoinsiemi di  $\Omega$  del tipo precedente siano eventi.

E' facile far vedere che, se si suppone che siano eventi i sottoinsiemi

$$\{\omega \in \Omega : x < X(\omega) \leq y\}$$

allora sono eventi tutti i sottoinsiemi del tipo

$$\{\omega \in \Omega : X(\omega) \in I\}$$

dove  $I$  è un qualsiasi intervallo (limitato o no, chiuso o no) di  $\mathbb{R}$ .

Intervalli di questo tipo sono ad esempio  $(-\infty, b)$ ,  $(-\infty, b]$ ,  $(a, b)$ ,  $(a, b]$ ,  $[a, b)$ ,  $[a, b]$ ,  $(a, +\infty)$  e  $[a, +\infty)$  con  $a, b \in \mathbb{R}$ .

Osserviamo che è del tipo considerato anche ogni sottoinsieme quale

$$\{\omega \in \Omega : X(\omega) = x\}$$

(basta prendere  $I = \{x\}$ , intervallo chiuso ridotto al solo punto  $x$ )

Più in generale si può far vedere che  $\{X \in I\}$  è un evento per una famiglia più vasta di sottoinsiemi  $I$  di  $\mathbb{R}$  (cioè non solo per gli intervalli). Tali insiemi sono detti **misurabili**. E' opportuno sapere che non tutti i sottoinsiemi di  $\mathbb{R}$  sono misurabili.

D'ora in avanti l'evento  $\{\omega \in \Omega : X(\omega) \in I\}$  sarà indicato con la notazione abbreviata  $\{X \in I\}$ . Allo stesso modo scriveremo  $\{x < X \leq y\}$  al posto di  $\{\omega \in \Omega : x < X(\omega) \leq y\}$ ,  $\{X \leq y\}$  invece di  $\{\omega \in \Omega : X(\omega) \leq y\}$  e così via.

Per evitare possibili equivoci, insistiamo sul fatto che quando si scrive, per esempio,  $\{1 \leq X \leq 2\}$ , non si intende l'intervallo  $[1, 2]$  ma un evento.

(1.2) Sia  $(\Omega, A, P)$  uno spazio di probabilità. Una **variabile aleatoria** (o anche **random variable**) è una funzione  $X : \Omega \rightarrow \mathbb{R}$  tale che, per ogni  $t \in \mathbb{R}$  risulti

$$\{X \leq t\} \in A$$

La precedente definizione è giustificata dal fatto che, affinché tutti i sottoinsiemi del tipo  $\{X \in I\}$ , con  $I$  insieme misurabile, siano eventi, è necessario e sufficiente supporre che lo siano anche i sottoinsiemi  $\{X \leq y\}$ .

Per brevità scriveremo semplicemente  $P(X \in I)$  anziché (come sarebbe corretto)  $P(\{X \in I\})$ .

(1.3) Sia  $X$  una variabile aleatoria su  $(\Omega, A, P)$ . Si chiama **legge** di  $X$  l'applicazione

$$I \rightarrow P(X \in I)$$

(definita sulla classe degli insiemi misurabili di  $\mathbb{R}$  e a valori in  $[0, 1]$ ).

La legge di una variabile aleatoria  $X$  va pensata come una “fotografia” delle varie probabilità assegnate a tutti gli eventi del tipo interessante (cioè  $\{X \in I\}$ ), e se ne intuisce dunque l'importanza. D'altra parte è evidente che nella pratica sarà impossibile “elencare” tutte le probabilità  $P(X \in I)$ , per ciascun  $I$ . Si capisce dunque la necessità di procurarsi alcuni strumenti non troppo complicati, che ci consentano di “risparmiare” calcoli.



## 2. Funzione di ripartizione di una variabile aleatoria: definizione e proprietà

(2.1) Sullo spazio di probabilità  $(\Omega, A, P)$  sia  $X$  una variabile aleatoria assegnata. Si chiama **funzione di ripartizione** (o **di distribuzione**) (f.d.r.) di  $X$  la funzione  $F: \mathbb{R} \rightarrow \mathbb{R}$  definita da

$$F(t) = P(X \leq t)$$

Quando saranno in gioco più variabili aleatorie  $X, Y, \dots$ , le rispettive funzioni di ripartizione verranno indicate con  $F_X, F_Y, \dots$ .

### (2.2) Teorema – Proprietà caratteristiche di una funzione di ripartizione

a) Siano  $X$  una variabile aleatoria,  $F$  la sua funzione di ripartizione. Valgono le seguenti proprietà:

- (i)  $0 \leq F(t) \leq 1, \quad \forall t \in \mathbb{R}$  (ovvio, in quanto  $F(t)$  rappresenta la probabilità di un evento, più precisamente dell'evento  $\{X \leq t\}$ )
- (ii)  $F$  è non decrescente, cioè, per ogni coppia  $x, y$  di numeri reali, con  $x \leq y$  risulta

$$F(x) \leq F(y)$$

Dimostrazione –  $F(x)$  rappresenta la probabilità dell'evento  $\{X \leq x\}$ , mentre  $F(y)$  rappresenta la probabilità dell'evento  $\{X \leq y\}$ . Essendo  $x \leq y$ , si deduce che  $\{X \leq x\} \subseteq \{X \leq y\}$  e quindi, per la proprietà di isotonia (la (2.4)<sup>1</sup>), si ha che  $F(x) \leq F(y)$

(iii) valgono le relazioni

$$\text{a) } \lim_{n \rightarrow -\infty} F(n) = 0 \qquad \text{b) } \lim_{n \rightarrow +\infty} F(n) = 1$$

(Dimostrazione di b) – Osserviamo che  $F(n)$  è la probabilità dell'evento  $A_n = \{X \leq n\}$ . La successione di eventi  $(A_n)_{n \in \mathbb{N}}$  è crescente ed inoltre  $\bigcup_{n \in \mathbb{N}} A_n = \Omega$ . Dunque, per la proprietà di passaggio al limite delle successioni crescenti di eventi (vedi punto (vi) – paragrafo 2)<sup>1</sup> si ha

$$\lim_{n \rightarrow +\infty} F(n) = \lim_{n \rightarrow +\infty} P(A_n) = P(\bigcup_{n \in \mathbb{N}} A_n) = P(\Omega) = 1$$

(Dimostrazione di a) – Osserviamo che  $F(n)$  è la probabilità dell'evento  $A_n = \{X \leq -n\}$ . La successione di eventi  $(A_n)_{n \in \mathbb{N}}$  è decrescente ed inoltre  $\bigcap_{n \in \mathbb{N}} A_n = \emptyset$ . Dunque, per la proprietà di passaggio al limite delle successioni decrescenti di eventi (vedi punto (vii) – paragrafo 2)<sup>1</sup> si ha

$$\lim_{n \rightarrow -\infty} F(n) = \lim_{n \rightarrow -\infty} P(A_n) = P(\bigcap_{n \in \mathbb{N}} A_n) = P(\emptyset) = 0$$

<sup>1</sup> – Riferimento al capitolo 1 (Spazi di probabilità)

(iv)  $F$  è continua a destra, cioè per ogni  $x \in \mathbb{R}$  si ha

$$F(x^+) = \lim_{n \rightarrow +\infty} F(x + \frac{1}{n}) = \lim_{t \rightarrow x^+} F(t) = F(x)$$

Per la definizione di funzione di ripartizione, possiamo scrivere

$$F(x^+) = \lim_{n \rightarrow +\infty} F(x + \frac{1}{n}) = \lim_{n \rightarrow +\infty} P(X \leq x + \frac{1}{n})$$

scomponiamo ora l'evento  $\{X \leq x + \frac{1}{n}\}$  in un'unione di due eventi disgiunti

$$\lim_{n \rightarrow +\infty} P(X \leq x + \frac{1}{n}) = \lim_{n \rightarrow +\infty} P\left((X \leq x) \cup (x < X \leq x + \frac{1}{n})\right)$$

applicando la proprietà di finita additività della probabilità, otteniamo

$$\lim_{n \rightarrow +\infty} P\left((X \leq x) \cup (x < X \leq x + \frac{1}{n})\right) = \lim_{n \rightarrow +\infty} \left[ P(X \leq x) + P(x < X \leq x + \frac{1}{n}) \right]$$

per le proprietà dei limiti si ricava che

$$\lim_{n \rightarrow +\infty} \left[ P(X \leq x) + P(x < X \leq x + \frac{1}{n}) \right] = \lim_{n \rightarrow +\infty} P(X \leq x) + \lim_{n \rightarrow +\infty} P(x < X \leq x + \frac{1}{n})$$

ove

$$\lim_{n \rightarrow +\infty} P(X \leq x) = \lim_{n \rightarrow +\infty} F(x) = F(x)$$

e

$\lim_{n \rightarrow +\infty} P(x < X \leq x + \frac{1}{n}) = 0$  in quanto la probabilità che dell'evento  $\{X \in (x, x + \frac{1}{n}]\}$  tende a 0 al crescere di  $n$

Per quanto evidente, si sottolinea il fatto che con la scrittura  $F(x^+)$  non si intende il valore di  $F$  nel punto (inesistente)  $x^+$ , ma il limite da destra verso  $x$ .

b) sia  $F : \mathbb{R} \rightarrow \mathbb{R}$  una funzione, che gode di tutte le proprietà (i), (ii), (iii), e (iv) elencate sopra. Allora esiste una variabile aleatoria  $X$  definita su un opportuno spazio di probabilità, che ammette  $F$  come funzione di ripartizione.

Il teorema precedente è utile quando si debba decidere se un'assegnata funzione  $F$  è una funzione di ripartizione: lo sarà se gode di tutte le proprietà (i)–(iv), non lo sarà se anche una sola di esse non è verificata.

(2.3) **Teorema** – La funzione di ripartizione individua (univocamente) la legge della variabile aleatoria. In altre parole se due variabili aleatorie hanno la stessa funzione di ripartizione, allora hanno la stessa legge (le due variabili si dicono *somiglianti* o anche *equamente distribuite*).

Vediamo in concreto il calcolo della legge di  $X$  a partire dalla sua funzione di ripartizione  $F$  in alcuni casi particolari (cioè il calcolo della quantità  $P(X \in I)$  per qualche tipo particolare di insieme misurabile  $I$ ).

(i)  $I = (a, b]$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

(ii)  $I = (-\infty, b)$

$$P(X < b) = \lim_{n \rightarrow +\infty} P(X \leq b - \frac{1}{n}) = \lim_{t \rightarrow b^-} F(t) = F(b^-)$$

con il simbolo  $F(b^-)$  si è indicato il limite verso  $b$  da sinistra, (e non il valore nel punto inesistente  $b^-$ ), e il calcolo precedente dice che tale limite esiste e vale  $P(X < b)$ .

(iii)  $I = \{x\}$

$$P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x^-)$$

(iv)  $I = [a, b)$

$$P(a \leq X < b) = P(X < b) - P(X < a) = F(b^-) - F(a^-)$$

(v)  $I = (a, b)$

$$P(a < X < b) = P(X < b) - P(X \leq a) = F(b^-) - F(a)$$

(vi)  $I = [a, b]$

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a^-)$$

(vii)  $I = (a, +\infty)$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$$

(viii)  $I = [a, +\infty)$

$$P(X \geq a) = 1 - P(X < a) = 1 - F(a^-)$$

La formula al punto (iii) è interessante. Ricordando che ogni funzione di ripartizione è continua a destra, essa può essere scritta anche nella forma equivalente

$$(2.4) \quad P(X = x) = F(x^+) - F(x^-)$$

ed interpretata dicendo che la probabilità che  $X$  assuma il valore  $x$  è esattamente uguale al “salto” che la funzione di ripartizione di  $X$  presenta nel punto di ascissa  $x$ .

I punti in cui  $F$  è continua sono tutti e soli i valori assunti dalla variabile  $X$  con probabilità nulla: questo segue immediatamente dalla relazione (2.4), osservando che  $F$  è continua in  $x$  se e soltanto se risulta  $F(x^+) - F(x^-) = 0$ .

Ciò giustifica la seguente definizione:

(2.5) La variabile aleatoria  $X$  è detta **continua** se la sua funzione di ripartizione  $F$  è continua, cioè se  $X$  assume con probabilità nulla ogni valore reale  $x$  (in simboli,  $P(X = x) = 0 \quad \forall x \in \mathbb{R}$  )

Esempio – La funzione di ripartizione definita da

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 & x \geq 1 \end{cases}$$

non è continua. In particolare, la variabile aleatoria  $X$  assume il solo valore 1 con probabilità uguale a 1 (il “salto della funzione di ripartizione”).

### 3. Variabili aleatorie discrete

(3.1) Sia  $X$  una variabile aleatoria definita sullo spazio  $(\Omega, A, P)$ .  $X$  si dice **discreta** se il suo dominio  $S$ , detto anche **spettro**, è un sottoinsieme finito o numerabile di  $\mathbb{R}$ .

(3.2) Sia  $X$  una variabile aleatoria discreta. L'applicazione  $p : \mathbb{R} \rightarrow \mathbb{R}$  definita da

$$p(x) = P(X = x)$$

si chiama **densità (discreta)** della variabile aleatoria  $X$ .

Quando saranno in gioco più variabili aleatorie  $X, Y, \dots$  useremo i simboli  $p_X, p_Y, \dots$  per indicare le rispettive densità.

(3.3) Sia  $p$  la densità di un'assegnata variabile aleatoria discreta  $X$ .  $p$  gode delle seguenti proprietà:

- (i)  $p(x) \geq 0 \quad \forall x \in \mathbb{R}$
- (ii)  $p(x) = 0$ , tranne quando  $x \in S$
- (iii)  $\sum_{x \in \mathbb{R}} p(x) = 1$

#### 4. Principali densità discrete

(4.1) **La densità binomiale** – Siano  $n \in \mathbb{N}$  e  $p \in (0, 1)$  due numeri fissati. Si chiama **densità binomiale** di parametri  $n$  e  $p$  (in simboli,  $B(n, p)$ ) la funzione così definita

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, 1, 2, \dots, n \\ 0 & \text{altrimenti} \end{cases}$$

Vediamo adesso una tipica situazione in cui compare la densità binomiale.

Sia  $(\Omega, A, P)$  lo schema delle prove indipendenti di parametri (noti)  $n$  e  $p$  (vedi 5.<sup>1</sup>).

Ricordiamo che ciò significa che

$$\Omega = \{0, 1\}^n$$

ed inoltre, se  $\omega = (\omega_1, \dots, \omega_n)$  è un evento elementare (con  $\omega_i \in \{0, 1\}$  per ogni  $i = 1, \dots, n$ ), si pone

$$P(\omega) = p^{(\sum_{i=1}^n \omega_i)} (1-p)^{(n - \sum_{i=1}^n \omega_i)}$$

Poniamo ora, per ogni  $i = 1, \dots, n$

$$X_i(\omega) = \omega_i$$

$$X(\omega) = \sum_{i=1}^n X_i(\omega) = \sum_{i=1}^n \omega_i$$

Sia dunque  $k$  un intero fissato, con  $k \in \{0, 1, 2, \dots, n\}$ . L'evento  $\{X = k\}$  è costituito dalle sequenze  $\omega$  che contengono il simbolo “1” esattamente  $k$  volte, cioè tali che

$$\sum_{i=1}^n \omega_i = k$$

Pertanto, per calcolare la probabilità di  $\{X = k\}$ , basterà sommare le probabilità di tutte le sequenze del tipo sopra detto, cioè

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

E' importante il caso particolare di densità  $B(1, p)$ ; la densità ottenuta si chiama anche **densità Bernoulliana**. Tale è la densità di una variabile aleatoria che assume i soli valori 1 e 0 con probabilità rispettive pari a  $p$  e  $1-p$ .

<sup>1</sup> – Riferimento al capitolo 1 (Spazi di probabilità)

(4.2) **La densità geometrica** – Sia  $p$  un numero assegnato, e supponiamo che sia  $0 < p < 1$ . Si chiama **densità geometrica** la funzione

$$p(k) = \begin{cases} (1-p)^{k-1} p & k = 1, 2, \dots \\ 0 & \text{altrove} \end{cases}$$

Una situazione in cui si presenta la densità geometrica è la seguente.

Supponiamo di eseguire una successione infinita di lanci di una stessa moneta, per la quale è  $p$  (assegnata) la probabilità di ottenere “1” in un singolo lancio. Ovviamente, al  $k$ -esimo lancio, avremo  $k-1$  “0” (la cui probabilità è  $1-p$ ) seguiti dal singolo simbolo “1” con  $k = 1, 2, \dots$ .

Da questo ricaviamo che

$$p(k) = (1-p)^{k-1} p \quad k = 1, 2, \dots$$

Nell’esperimento degli infiniti lanci di una moneta, potrebbe verificarsi l’evento

B: “il simbolo “1” non esce mai”. Ma con quale probabilità avviene ciò?

Iniziamo a porre

$A_k$  : “esce “1” al  $k$ -esimo lancio”

Si può scrivere ora

$$B^c = \bigcup_{k=1}^{\infty} A_k$$

poiché gli eventi  $A_k$  sono due a due disgiunti, per la proprietà di  $\sigma$ -additività di  $P$  si ha

$$P(B) = 1 - P\left(\bigcup_{k=1}^{\infty} A_k\right) = 1 - \sum_{k=1}^{\infty} P(A_k) = 1 - \sum_{k=1}^{\infty} (1-p)^{k-1} p$$

ricordando la somma della serie geometrica, abbiamo

$$1 - \sum_{k=1}^{\infty} (1-p)^{k-1} p = 1 - p \frac{1}{1-(1-p)} = 0$$

Cioè: per quanto piccola (purché strettamente positiva) sia la probabilità  $p$  di ottenere “1” in un singolo lancio, l’eventualità che “1” non esca mai è “quasi impossibile” (cioè di probabilità nulla).

(4.3) **La densità di Poisson** – Sia  $\lambda > 0$  un numero reale assegnato.

La funzione

$$p(k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & k = 0, 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases}$$

è una densità. Essa va sotto il nome di **densità di Poisson** di parametro  $\lambda$ .

Per ottenere in modo naturale questa densità, consideriamo la seguente situazione:

per ogni intero  $n$ , sia  $X_n$  una variabile aleatoria avente densità  $B(n, p = \lambda/n)$ , cioè

$$P(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

Ci interessa trovare il limite di questa probabilità al tendere di  $n$  all'infinito (cioè, ci interessa trovare a quali numeri si avvicinano i valori della densità binomiale quando il numero  $n$  di prove ripetute aumenta e la probabilità  $\lambda/n$  di successo in una singola prova diminuisce).

Fissato ora  $k$ , sia  $n$  abbastanza grande in modo che sia  $n \geq k$ . Si ha

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} = \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{n!}{(n-k)! n^k} \end{aligned}$$

ricordando, dalla (6)<sup>1</sup>, che  $\frac{n!}{(n-k)!} = (n)_k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$ , abbiamo

$$P(X_n = k) = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{n!}{(n-k)! n^k} = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \underbrace{\frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n}}_{k \text{ fattori}}$$

Osservando che

- (i)  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$
- (ii)  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$
- (iii)  $\lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} = 1$

se ne deduce che

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

cioè la densità di Poisson di parametro  $\lambda$ , calcolata in  $k$ .

<sup>1</sup> – Riferimento al capitolo 2 (Richiami di calcolo combinatorio)

(4.4) **La densità ipergeometrica** – Assegnati tre numeri interi strettamente positivi  $a$ ,  $b$  e  $n$ , con  $n \leq a + b$ , poniamo  $k_1 = \max(0, n - b)$  e  $k_2 = \min(a, n)$  e osserviamo che è  $k_1 \leq k_2$ .

La funzione definita nel modo seguente

$$p(k) = \begin{cases} \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}} & k = k_1, k_1 + 1, \dots, k_2 \\ 0 & \text{altrove} \end{cases}$$

è una densità discreta. Essa prende il nome di **densità ipergeometrica** di parametri  $a$ ,  $b$  e  $n$ .

Vediamo una situazione in cui questa densità si presenta in modo naturale.

Supponiamo di avere un'urna contenente  $a$  palline rosse e  $b$  palline verdi. Si eseguono  $n$  estrazioni senza rimpiazzo e poniamo  $X$  = numero di palline rosse (tra le  $n$  estratte).

Costruiamo uno spazio di probabilità  $(\Omega, A, P)$  con cui modellizzare l'esperimento: ogni possibile risultato è un gruppo di  $n$  palline prese dalle  $a + b$  palline presenti nell'urna; quindi l'insieme  $\Omega$  degli eventi può essere identificato con l'insieme delle combinazioni di  $a + b$  oggetti a  $n$  a  $n$  (vedi (10)<sup>1</sup>). Inoltre, poiché non c'è motivo di pensare che un particolare risultato sia più o meno probabile di ogni altro, assegniamo a ciascun  $\omega \in \Omega$  la stessa probabilità, pari a

$$P(\omega) = \frac{1}{\text{card}\Omega} = \frac{1}{\binom{a+b}{n}}$$

Premesso ciò, è immediato verificare che sullo spazio così costruito  $X$  è una variabile aleatoria, e ciò che resta da fare è calcolarne la densità, cioè le quantità  $P(X = k)$  per ogni intero  $k$ .

E' chiaro che  $P(X = k)$  sarà diversa da zero solo per gli interi  $k$  compresi tra  $k_1$  e  $k_2$  (infatti: il numero di palline rosse estratte non può superare né il numero di palline rosse a disposizione ( $= a$ ), né il numero totale di estrazioni effettuate ( $= n$ ). Inoltre, se  $n \geq b$ ,  $n - b$  è il minimo numero di palline rosse che possono essere estratte).

L'evento  $\{X = k\}$  corrisponde al sottoinsieme di  $\Omega$  costituito dai gruppi di  $n$  palline delle quali  $k$  sono rosse e  $n - k$  verdi, e dunque ha cardinalità pari a

$$\binom{a}{k} \binom{b}{n-k}$$

Dato che stiamo lavorando su uno spazio uniforme si conclude che

$$P(X = k) = \frac{\text{card}\{X = k\}}{\text{card}\Omega} = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}$$

<sup>1</sup> – Riferimento al capitolo 2 (Richiami di calcolo combinatorio)



## 5. Vettori aleatori discreti

Può capitare che l'esperimento sotto osservazione produce più variabili interessanti, e che di esse si debba studiare il comportamento "congiunto": ad esempio, in uno schema di 5 prove indipendenti, consideriamo le due quantità  $X$  = numero di successi nelle 5 prove e  $Y$  = numero di successi nelle prime due prove; in tal caso una domanda interessante a cui rispondere potrebbe essere: quanto vale la probabilità che su un totale di 3 successi, due di essi si siano presentati nelle prime due prove? In simboli, questo significa calcolare la probabilità congiunta

$$P(\{X = 3\} \cap \{Y = 2\})$$

Supporremo fissato una volta per tutte lo spazio di probabilità  $(\Omega, A, P)$ .

(5.1) Siano  $X_1, X_2, \dots, X_m$   $m$  variabili aleatorie discrete (tutte definite su  $(\Omega, A, P)$ ). Si chiama **vettore aleatorie discreto**  $m$ -dimensionale (su  $(\Omega, A, P)$ ) la  $m$ -upla di variabili aleatorie

$$X = (X_1, X_2, \dots, X_m)$$

si tratta della funzione  $X : \Omega \rightarrow \mathbb{R}^m$  definita da

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_m(\omega))$$

L'attributo *aleatorio* dato al vettore  $X$  è giustificato dal fatto seguente: per ogni  $i = 1, 2, \dots, m$  sia  $x_i \in \mathbb{R}$ . Come sappiamo, il sottoinsieme di  $\Omega$   $\{X_i = x_i\}$  è un evento, dunque, per la proprietà di stabilità della  $\sigma$ -algebra  $A$  rispetto all'intersezione finita, è un evento anche il sottoinsieme

$$\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \dots \cap \{X_m = x_m\}$$

che viene indicato più brevemente con la scrittura

$$\{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\}$$

Ha quindi senso considerare la probabilità degli eventi del tipo precedente e si ha la seguente definizione

(5.2) Indichiamo con  $x = (x_1, x_2, \dots, x_m)$  il generico punto di  $\mathbb{R}^m$ . Si chiama **densità** del vettore  $X$  (o **densità congiunta** delle variabili  $X_1, X_2, \dots, X_m$ ) la funzione  $p : \mathbb{R}^m \rightarrow \mathbb{R}$  definita da

$$p(x) = p(x_1, x_2, \dots, x_m) = P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m)$$

Riprendiamo ora l'esempio citato all'inizio del paragrafo.

Siano  $S_X = \{0, 1, 2, 3, 4, 5\}$  e  $S_Y = \{0, 1, 2\}$  i rispettivi spettri delle due variabili aleatorie.

La seguente matrice, in cui ogni  $p(i, j) = P(X = i, Y = j)$ , è detta **matrice delle probabilità congiunte**.

$X / Y$	0	1	2
0	$p(0,0)$	$p(0,1)$	$p(0,2)$
1	$p(1,0)$	$p(1,1)$	$p(1,2)$
2	$p(2,0)$	$p(2,1)$	$p(2,2)$
3	$p(3,0)$	$p(3,1)$	$p(3,2)$
4	$p(4,0)$	$p(4,1)$	$p(4,2)$
5	$p(5,0)$	$p(5,1)$	$p(5,2)$

Le quantità del tipo

$$p(., j) = \sum_{i=0}^5 p(i, j) = P(Y = j), \quad p(i, .) = \sum_{j=0}^2 p(i, j) = P(X = i)$$

vengono chiamate anche **densità marginali**.

Da ciò ricaviamo che è possibile ottenere le densità marginali quando si conosce la densità congiunta (mentre, in generale, le densità marginali non individuano in modo univoco la congiunta). Inoltre,

$$\sum_{i=0}^5 \sum_{j=0}^2 p(i, j) = \sum_{i=0}^5 p(i, .) = \sum_{i=0}^5 P(X = i) = 1$$

$$\sum_{j=0}^2 \sum_{i=0}^5 p(i, j) = \sum_{j=0}^2 p(., j) = \sum_{j=0}^2 P(Y = j) = 1$$

cioè, la somma delle componenti della matrice è pari a 1.

## 6. Variabili aleatorie indipendenti

Supporremo al solito fissato lo spazio di probabilità  $(\Omega, A, P)$ . Può accadere che due variabili aleatorie  $X$  e  $Y$  siano tali che il sapere  $X$  assume un valore in un certo insieme  $A$  non cambi la valutazione di probabilità che  $Y$  assuma un valore in un dato insieme  $B$ : in altre parole, tali che gli eventi  $\{X \in A\}$  e  $\{Y \in B\}$  siano indipendenti, e ciò per ogni coppia di sottoinsiemi  $A$  e  $B$  di  $R$ .

Due variabili aleatorie  $X$  e  $Y$  si dicono tra loro **indipendenti** se, per ogni coppia di sottoinsiemi  $A$  e  $B$  di  $R$ , risulta

$$(6.1) \quad P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Più in generale, le variabili aleatorie  $X_1, X_2, \dots, X_m$  si dicono tra loro **indipendenti** se, per ogni famiglia  $A_1, A_2, \dots, A_m$  di sottoinsiemi di  $R$ , risulta

$$(6.2) \quad P(X_1 \in A_1, X_2 \in A_2, \dots, X_m \in A_m) = \prod_{i=1}^m P(X_i \in A_i)$$

Siano  $X_1, X_2, \dots, X_m$   $m$  variabili aleatorie discrete, aventi densità congiunta  $p(x_1, x_2, \dots, x_m)$  e densità marginali  $p_{X_1}(x_1), p_{X_2}(x_2), \dots, p_{X_m}(x_m)$ . Allora  $X_1, X_2, \dots, X_m$  sono indipendenti tra loro se e solo se,  $\forall (x_1, x_2, \dots, x_m) \in R^m$  vale la relazione

$$(6.3) \quad p_{X_1}(x_1), p_{X_2}(x_2), \dots, p_{X_m}(x_m) = p_{X_1}(x_1) \times p_{X_2}(x_2) \times \dots \times p_{X_m}(x_m)$$

Un'ultima definizione importante è la seguente:

Siano  $X$  e  $Y$  due variabili aleatorie discrete, aventi densità congiunta  $p(x, y)$ . Si chiama **densità condizionale** di  $Y$ , dato  $X = x$ , la funzione definita come segue

$$y \mapsto p_{Y|X}(y|x) = \begin{cases} \frac{P(Y=y, X=x)}{P(X=x)} = \frac{p(x, y)}{p_X(x)} & \text{se } p_X(x) \neq 0 \\ 0 & \text{altrimenti} \end{cases}$$

## 7. Riepilogo delle principali densità discrete

Nome	Densità	Legge	Situazione in cui si applica (breve descrizione)
Binomiale	$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$	$X \sim B(n, p)$	Esperimento di $n$ prove ripetute (indipendenti); $p$ rappresenta la probabilità di successo in una singola prova
Bernoulliana	$p(k) = p^k (1-p)^{1-k} \quad k = 0, 1$	$X \sim B(1, p)$	Caso particolare della binomiale in cui il numero di prove è pari a 1
Geometrica	$p(k) = (1-p)^{k-1} p \quad k = 1, 2, \dots$	$X \sim G(p)$	Esperimento di prove (indipendenti) ripetute fino ad ottenere il successo; $p$ rappresenta la probabilità di successo in una singola prova
Poisson	$p(k) = \frac{a^k}{k!} e^{-a} \quad k = 0, 1, \dots, n$	$X \sim \text{Poisson}(a)$	Caso particolare della binomiale in cui: $n$ , il numero di prove, viene fatto tendere a $+\infty$ ; $p$ , la probabilità di successo di una singola prova, a 0; $a$ , il parametro della densità, è pari a $np$
Ipergeometrica	$p(k) = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}$ $k = \max\{0, n-b\}, \dots, \min\{a, n\}$	$X \sim IG(a, b, n)$	Esperimento di $n$ prove ripetute (non indipendenti);

# Variabili aleatorie assolutamente continue

## 1. Definizioni e prime proprietà

Abbiamo visto che, se  $X$  è una variabile aleatoria discreta di densità  $p$ , allora, per ogni  $A \subseteq \mathbb{R}$ , si ha

$$(1.1) \quad P(X \in A) = \sum_{x \in A} p(x)$$

Una variabile aleatoria sullo spazio  $(\Omega, A, P)$  si dice **assolutamente continua** se esiste una funzione  $f: \mathbb{R} \rightarrow \mathbb{R}^+$  integrabile su  $\mathbb{R}$  e tale che, per ogni insieme misurabile  $A$  di  $\mathbb{R}$ , risulta

$$(1.2) \quad P(X \in A) = \int_A f(x) dx$$

In tal caso la funzione  $f$  si chiama **densità** di  $X$ .

Come al solito, quando saranno in gioco più variabili  $X, Y, \dots$ , indicheremo le rispettive densità con  $f_X, f_Y, \dots$ .

Si noti l'analogia tra le formule (1.1) e (1.2). La seconda si può formalmente ottenere dalla prima sostituendo il simbolo di sommatoria con quello di integrale.

Dalla formula (1.2) si deduce in particolare che la funzione di ripartizione  $F$  di  $X$  assuma la forma

$$F(t) = P(X \leq t) = P(X \in (-\infty, t]) = \int_{-\infty}^t f(x) dx$$

Inoltre si ha

$$P(a < X \leq b) = P(X \in (a, b]) = \int_a^b f(x) dx$$

E ancora

$$P(X \in \{a\}) = \int_{\{a\}} f(x) dx = 0$$

da cui segue che la variabile aleatoria  $X$  è continua.

Infine si ha (notare l'analogia con il caso discreto)

$$\int_{-\infty}^{+\infty} f(x) dx = P(X \in \mathbb{R}) = P(\Omega) = 1$$

Due variabili aleatorie si dicono tra loro indipendenti se per ogni coppia  $A, B$  di sottoinsiemi misurabili di  $\mathbb{R}$ , risulta

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

oppure se, per ogni coppia di numeri reali  $s, t$ , risulta

$$P(X \leq s, Y \leq t) = P(X \leq s)P(Y \leq t) = F_X(s)F_Y(t)$$

Sia  $X$  una variabile aleatoria assolutamente continua avente densità  $f$ .

(1.3) Si dice che  $X$  ha *speranza matematica finita* se

$$\int_{-\infty}^{+\infty} |x| f(x) dx < \infty$$

in tal caso si chiama *speranza matematica* di  $X$  il numero

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

(1.4) Se  $X$  ha speranza  $\mathbf{E}[X]$  finita e se anche

$$\int_{-\infty}^{+\infty} (x - \mathbf{E}[X])^2 f(x) dx < \infty$$

si dice che  $X$  ha *varianza finita* e si pone

$$\mathbf{Var}X = \mathbf{E}[(X - \mathbf{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathbf{E}[X])^2 f(x) dx$$

Per la speranza e la varianza delle variabili aleatorie assolutamente continue valgono tutte le proprietà che abbiamo visto nel caso discreto. Si noti l'analogia delle definizioni: il simbolo di sommatoria che usavamo nel caso discreto è stato sostituito dal simbolo di integrale.

## 2. Alcune densità di variabili aleatorie assolutamente continue

(2.1) **Le leggi uniformi** – Si sceglie un punto a caso nell'intervallo  $[a, b]$  della retta  $\mathbb{R}$ . Indichiamo con  $X$  l'ascissa del punto scelto. La funzione di ripartizione di  $X$  è

$$F(t) = \begin{cases} 0 & \text{se } t < a \\ \frac{t-a}{b-a} & \text{se } a \leq t \leq b \\ 1 & \text{se } t > b \end{cases}$$

e la densità è

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$$

La legge qui descritta si chiama **legge uniforme** (sull'intervallo  $[a, b]$  e si indica con il simbolo  $U(a, b)$ ); ogni variabile aleatoria che ammetta questa legge si dice **uniforme** (su  $[a, b]$ ).

(2.1.1) Calcolo della speranza matematica di  $X$ :

Per la definizione di speranza abbiamo

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \frac{1}{b-a} dx$$

da cui si ha

$$\mathbf{E}[X] = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

Si noti che  $\int_{-\infty}^{+\infty}$  diventa semplicemente  $\int_a^b$  in quanto non vengono considerati quei punti non appartenenti all'intervallo  $[a, b]$ , per i quali la  $f$  risulta pari a 0.

(2.1.2) Calcolo della varianza di  $X$ :

Per la (1.11)<sup>1</sup> abbiamo

$$\mathbf{Var}X = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

cominciando dal calcolo di  $\mathbf{E}[X^2]$

$$\mathbf{E}[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(a^2 + b^2 + ab)}{3(b-a)} = \frac{a^2 + b^2 + ab}{3}$$

Di conseguenza

$$\mathbf{Var}X = \frac{a^2 + b^2 + ab}{3} - \frac{a^2 + b^2 + 2ab}{4} = \frac{a^2 + b^2 - 2ab}{12} = \frac{(b-a)^2}{12}$$

Da ciò segue

$$\sqrt{\mathbf{Var}X} = \frac{b-a}{2\sqrt{3}}$$

(2.1.3) Consideriamo il caso in cui  $a=0$  e  $b=1$ , cioè  $X \sim U(0,1)$ .

Per quello detto finora, avremo

$$F(t) = \begin{cases} 0 & \text{se } t < 0 \\ t & \text{se } 0 \leq t \leq 1 \\ 1 & \text{se } t > 1 \end{cases}$$

$$f(x) = \begin{cases} 1 & \text{se } 0 \leq x \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

$$\mathbf{E}[X] = \frac{1}{2} \text{ e } \mathbf{Var}X = \frac{1}{12}$$

<sup>1</sup> – Riferimento al capitolo 4 (Speranza matematica e varianza)



(2.2) **La legge esponenziale** – Sia  $\lambda$  un numero reale positivo assegnato,  $X$  una variabile aleatoria. Si dice che  $X$  ha **legge esponenziale** (di parametro  $\lambda$ ) (il simbolo è  $\mathcal{E}(\lambda)$  oppure  $Esp(\lambda)$ ) se ammette come densità la funzione

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \lambda e^{-\lambda x} & \text{se } x \geq 0 \end{cases}$$

La funzione di ripartizione è

$$F(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ 1 - e^{-\lambda t} & \text{per } t > 0 \end{cases}$$

(2.2.1) Calcolo della speranza matematica di  $X$ :

Per la definizione di speranza abbiamo

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \lambda e^{-\lambda x} dx$$

da cui si ha

$$\begin{aligned} \mathbf{E}[X] &= \int_0^{+\infty} x \lambda e^{-\lambda x} dx = x - e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= 0 + \frac{1}{\lambda} \cdot \int_0^{+\infty} \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \cdot 1 = \frac{1}{\lambda} \end{aligned}$$

Si noti che  $\int_{-\infty}^{+\infty}$  diventa semplicemente  $\int_0^{+\infty}$  in quanto non vengono considerati quei punti non appartenenti all'intervallo  $[0, +\infty)$ , per i quali la  $f$  risulta pari a 0.

(2.2.2) Calcolo della varianza di  $X$ :

Come al solito abbiamo  $\mathbf{Var}X = \mathbf{E}[X^2] - \mathbf{E}[X]^2$ ; cominciando dal calcolo di  $\mathbf{E}[X^2]$

$$\begin{aligned} \mathbf{E}[X^2] &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = x^2 - e^{-\lambda x} \Big|_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx \\ &= 0 + \frac{1}{\lambda} \cdot 2 \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \cdot \frac{2}{\lambda} = \frac{2}{\lambda^2} \end{aligned}$$

Di conseguenza

$$\mathbf{Var}X = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Da ciò segue

$$\sqrt{\mathbf{Var}X} = \frac{1}{\lambda} = \mathbf{E}[X]$$

Ogni variabile  $X$  avente legge esponenziale verifica la seguente relazione

$$(2.2.3) \quad P(X > t + s | X > t) = P(X > s)$$

dove  $s$  e  $t$  sono due numeri positivi assegnati.

Dimostrazione – Per la definizione di probabilità condizionata abbiamo

$$P(X > t + s | X > t) = \frac{P(X > t + s, X > t)}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)}$$

dove, per la definizione di densità di una variabile aleatoria assolutamente continua, vale

$$\frac{P(X > t + s)}{P(X > t)} = \frac{\int_{t+s}^{+\infty} \lambda e^{-\lambda x} dx}{\int_t^{+\infty} \lambda e^{-\lambda x} dx}$$

da cui si ha

$$\frac{P(X > t + s)}{P(X > t)} = \frac{\int_{t+s}^{+\infty} \lambda e^{-\lambda x} dx}{\int_t^{+\infty} \lambda e^{-\lambda x} dx} = \frac{-e^{-\lambda x} \Big|_{t+s}^{+\infty}}{-e^{-\lambda x} \Big|_t^{+\infty}} = \frac{-e^{-\lambda(t+s)}}{-e^{-\lambda t}} = \frac{e^{-\lambda t} \cdot e^{-\lambda s}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$$

La relazione (2.2.3) viene detta **proprietà di assenza di memoria** o anche **proprietà di assenza di usura**. Per capire il significato di queste espressioni, immaginiamo che  $X$  rappresenti la durata di una lampadina, messa in funzione all'istante 0. Allora la (2.2.3) si interpreta così: sapendo che la lampadina non si è guastata fino all'istante  $t$ , la probabilità che essa duri per ulteriori  $s$  istanti è uguale alla probabilità che la lampadina messa in funzione all'istante 0, non si guasti fino all'istante  $s$ . In altri termini, la lampadina non è soggetta all'usura causata dai primi  $t$  istanti di funzionamento.

(2.3) **La legge di Weibull** – Siano  $\alpha > 0$  e  $\lambda > 0$  due numeri fissati, e consideriamo la funzione

$$f(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

Osserviamo anche che, per  $\alpha = 1$ , si ottiene la densità esponenziale di parametro  $\lambda$  di cui abbiamo già parlato ( $\mathcal{E}(\lambda) = W(\lambda, 1)$  oppure  $\text{Esp}(\lambda) = W(\lambda, 1)$ ). Nel caso generale,  $f$  viene chiamata **densità di Weibull** (di parametri  $\alpha$  e  $\lambda$ , in simboli  $W(\lambda, \alpha)$ ).

Sia  $X$  una variabile aleatoria avente  $f$  come densità. Ci interessa studiare la quantità  $P(X > t + s | X > t)$ . In questo caso si ha

$$P(X > t + s | X > t) = \frac{\int_t^{t+s} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} dx}{\int_t^{+\infty} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} dx} = \frac{-e^{-\lambda x^\alpha} \Big|_t^{t+s}}{-e^{-\lambda x^\alpha} \Big|_t^{+\infty}} = \frac{-e^{-\lambda (t+s)^\alpha}}{-e^{-\lambda t^\alpha}} = e^{-\lambda [(t+s)^\alpha - t^\alpha]}$$

Per ogni  $s$  fissato, la funzione  $t \mapsto P(X > t + s | X > t)$  è crescente (risp. costante, decrescente) quando la funzione  $t \mapsto (t + s)^\alpha - t^\alpha$  è decrescente (risp. costante, crescente). Ciò accade quando

$$\frac{d}{ds} [(t + s)^\alpha - t^\alpha] = \alpha [(t + s)^{\alpha-1} - t^{\alpha-1}] < 0$$

(risp.  $= 0, > 0$ ) ovvero quando  $\alpha < 1$  (risp.  $= 1, > 1$ ). Ricapitolando

$$t \mapsto P(X > t + s | X > t) \begin{cases} \text{crescente per } \alpha < 1 \\ \text{costante per } \alpha = 1 \\ \text{decrescente per } \alpha > 1 \end{cases}$$

Supponiamo che  $X$  rappresenti la durata di una lampadina soggetta ad usura. E' naturale allora scegliere per  $X$  una densità per la quale la funzione  $t \mapsto P(X > t + s | X > t)$  sia decrescente, e per tale motivo in questo caso viene usata una variabile  $X$  avente densità di Weibull di parametro  $\alpha > 1$ . Una situazione in cui sarebbe invece indicata una densità di Weibull con parametro  $\alpha < 1$  è ad esempio quella in cui  $X$  rappresenta la durata di vita di un neonato (è noto che la probabilità di mortalità infantile è massima nei primi giorni di vita del bambino, per poi andare a diminuire man mano che i giorni passano).

### 3. Le principali densità della statistica

(3.1) **La densità normale** – Siano  $\mu \in \mathbb{R}$  e  $\sigma \in \mathbb{R}^+$  due parametri assegnati. Si chiama **densità normale** (o **gaussiana**) la funzione

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$$

La legge di cui  $f$  è una densità si indica con il simbolo  $N(\mu, \sigma^2)$ ; ogni variabile aleatoria avente tale legge si chiama variabile aleatoria **normale** o **gaussiana**.

Nella famiglia delle leggi normali, sopra definite, particolare rilevanza ha la  $N(0, 1)$  (cioè la legge di parametri  $\mu = 0$  e  $\sigma^2 = 1$ ). Scriviamone esplicitamente la densità:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Appunto per la sua importanza, questa legge viene detta **normale standard** (o **gaussiana standard**). Una prima proprietà della famiglia delle leggi normali è la sua stabilità rispetto alle trasformazioni affini; precisamente si ha:

(3.1.1) **Proposizione** – Sia  $X$  una variabile aleatoria avente legge gaussiana  $N(m, s^2)$ ; siano inoltre  $a$  e  $b$  due numeri reali. Allora la variabile aleatoria  $Y = aX + b$  ha legge  $N(am + b, a^2 s^2)$ .

#### Dimostrazione

(i) Per  $a > 0$

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) = P(aX \leq y - b) = P\left(X \leq \frac{y-b}{a}\right) \\ P\left(X \leq \frac{y-b}{a}\right) &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}s} e^{-\frac{(t-m)^2}{2s^2}} dt = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}as} e^{-\frac{(x-am-b)^2}{2a^2s^2}} dx \end{aligned}$$

dove l'ultima eguaglianza si ottiene effettuando il cambiamento di variabile

$$t = \frac{x-b}{a}$$

(ii) Per  $a < 0$

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) = P(aX \leq y - b) = P\left(X \geq \frac{y-b}{a}\right) \\ P\left(X \geq \frac{y-b}{a}\right) &= \int_{\frac{y-b}{a}}^{+\infty} \frac{1}{\sqrt{2\pi}s} e^{-\frac{(t-m)^2}{2s^2}} dt = \int_y^{+\infty} \frac{1}{\sqrt{2\pi}as} e^{-\frac{(x-am-b)^2}{2a^2s^2}} dx = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}as} e^{-\frac{(x-am-b)^2}{2a^2s^2}} dx \end{aligned}$$

Dalla (3.1.1) si deduce il

(3.1.2) **Corollario** – Siano  $\mu \in \mathbb{R}$  e  $\sigma \in \mathbb{R}^+$  due numeri assegnati.

(i) Se  $X \sim N(0, 1)$ , allora  $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$

(ii) Se  $X \sim N(\mu, \sigma^2)$ , allora  $Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$

Calcoliamo ora la speranza e la varianza di una variabile aleatoria  $X$  avente legge  $N(\mu, \sigma^2)$ , cominciamo dal caso particolare  $\mu = 0$  e  $\sigma^2 = 1$ .

Sia dunque  $X \sim N(0, 1)$ .

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} = 0$$

Passiamo ora al calcolo della varianza (che, poiché  $\mathbf{E}[X] = 0$ , in questo caso coincide con  $\mathbf{E}[X^2]$ ).

$$\begin{aligned} \mathbf{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} x \left( x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right) dx = \int_{-\infty}^{+\infty} x \frac{d}{dx} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right) dx \\ &= -x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 \end{aligned}$$

A questo punto siamo in grado di calcolare  $\mathbf{E}[X]$  e  $\mathbf{Var}X$  anche per una variabile aleatoria  $X$  avente legge  $N(\mu, \sigma^2)$ . Per far questo usiamo il punto (ii) del corollario (3.1.2), che ci assicura che la variabile aleatoria definita da

$$Y = \frac{X - \mu}{\sigma}$$

ha legge  $N(0, 1)$ .

D'altra parte si può scrivere

$$X = \sigma Y + \mu$$

e quindi, per le note proprietà della speranza e della varianza, risulta

$$\mathbf{E}[X] = \mathbf{E}[\sigma Y + \mu] = \sigma \mathbf{E}[Y] + \mu = \mu$$

$$\mathbf{Var}X = \mathbf{Var}(\sigma Y + \mu) = \sigma^2 \mathbf{Var}(Y) = \sigma^2$$

(3.1.3) **Teorema** – Siano  $X_1, X_2, \dots, X_n$   $n$  variabili aleatorie indipendenti, tali che, per ogni  $i$ , con  $i = 1, 2, \dots, n$  si abbia

$$X_i \sim N(\mu_i, \sigma_i^2)$$

Allora la variabile aleatoria

$$Z = X_1 + X_2 + \dots + X_n$$

ha legge  $N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$ .

(3.1.4) Siano  $X_1, X_2, \dots, X_n$   $n$  variabili aleatorie indipendenti e tutte con la stessa legge (non necessariamente gaussiana). Poniamo

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

La variabile aleatoria  $\bar{X}$  viene detta **media campionaria** ed è di uso comune in statistica.

Calcoleremo ora la legge di  $\bar{X}$  nel caso particolare in cui le  $X_i, i = 1, 2, \dots, n$  abbiano legge  $N(\mu, \sigma^2)$ . Posto  $Z = X_1 + X_2 + \dots + X_n$ , per il teorema precedente si ha

$$Z \sim N(n\mu, n\sigma^2)$$

dunque, per la (3.1.1) applicata a  $Z$  (con valori  $a = \frac{1}{n}$  e  $b = 0$ ) si ha

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Come conseguenza, per il punto (ii) di (3.1.2) si deduce che la variabile aleatoria centrata e ridotta

$$\begin{aligned} \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} &= \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} = \sqrt{n} \frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\sigma} = \frac{\sqrt{n}}{\sqrt{n}} \sqrt{n} \frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\sigma} \\ &= n \frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\sigma \sqrt{n}} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma \sqrt{n}} \end{aligned}$$

ha legge normale standard.

D'ora in avanti la funzione di ripartizione della legge  $N(0, 1)$  verrà indicata con il simbolo  $\Phi$ ; abbiamo cioè

$$\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad t \in \mathbb{R}$$

Abbiamo visto che, se si hanno  $n$  variabili indipendenti  $X_1, X_2, \dots, X_n$  tutte con legge  $N(\mu, \sigma^2)$  e a partire da esse si costruisce la variabile aleatoria centrata e ridotta associata alla loro media campionaria, cioè

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

la variabile  $S_n^*$  risulta avere legge normale standard.

(3.1.5) **Teorema limite centrale** (abbreviato con la sigla **CLT** nella letteratura di lingua anglosassone) – Sia  $(X_n)_{n \in \mathbb{N}}$  una successione di variabili aleatorie indipendenti e tutte con la stessa legge (si abbrevia con la sigla: i.i.d. = indipendenti identicamente distribuite), aventi media  $\mu$  e varianza  $\sigma^2$  finite. Per ogni intero  $n$  poniamo

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

Allora, per ogni  $x \in \mathbb{R}$  si ha

$$\lim_{n \rightarrow \infty} P(S_n^* \leq x) = \Phi(x)$$

Si ha cioè il fatto che la legge di  $S_n^*$  tende ad avvicinarsi alla legge normale standard qualunque sia la legge di partenza delle variabili  $(X_i)$  (non necessariamente gaussiane).

In genere basta  $n > 50$  per avere un'approssimazione soddisfacente.

(3.2) **La legge dei  $3\sigma$**  – Sia  $X$  una variabile aleatoria con legge  $N(\mu, \sigma^2)$ . Cerchiamo di calcolare le seguenti probabilità:

- (i)  $P(\mu - \sigma^2 \leq X \leq \mu + \sigma^2)$
- (ii)  $P(\mu - 2\sigma^2 \leq X \leq \mu + 2\sigma^2)$
- (iii)  $P(\mu - 3\sigma^2 \leq X \leq \mu + 3\sigma^2)$

Passando alla variabile aleatoria normale standard  $Y = \frac{X - \mu}{\sigma}$ , abbiamo

- (i)  $P(\mu - \sigma^2 \leq X \leq \mu + \sigma^2) = P(-1 \leq Y \leq 1) = 0.6826$
- (ii)  $P(\mu - 2\sigma^2 \leq X \leq \mu + 2\sigma^2) = P(-2 \leq Y \leq 2) = 0.9544$
- (iii)  $P(\mu - 3\sigma^2 \leq X \leq \mu + 3\sigma^2) = P(-3 \leq Y \leq 3) = 0.9974$

Da questo si nota che è minore dello 0.3% la probabilità, che una variabile aleatoria con legge  $N(\mu, \sigma^2)$ , assuma un valore che cada al di fuori dell'intervallo semidimensionale  $3\sigma$  avente centro in  $\mu$ . Questa è la cosiddetta “**legge dei  $3\sigma$** ” secondo cui è altamente improbabile che una variabile aleatoria normale assuma valori che distano dalla sua media per più di tre deviazioni standard.



#### 4. La funzione di ripartizione della legge normale standard

(4.1) Sia  $X$  una variabile aleatoria. Si dice che  $X$  è simmetrica se  $X$  e  $-X$  hanno la stessa legge, o, in modo equivalente, la stessa funzione di ripartizione:

$$F_X(t) = P(X \leq t) = P(-X \leq t) = F_{-X}(t)$$

La precedente relazione si può scrivere anche nella forma

$$P(X \leq -t) = P(X \geq t) = 1 - P(X \leq t)$$

ovvero anche

$$(4.1.1) \quad P(X \leq t) + P(X \geq t) = F_X(t) + F_X(-t) = 1$$

Dalla relazione precedente si deduce in particolare che, se  $X$  è simmetrica allora  $2F_X(0) = 1$ , e cioè

$$(4.1.2) \quad F_X(0) = \frac{1}{2}$$

Se  $X$  ammette densità  $f$ , allora  $X$  è simmetrica se e solo se,  $\forall t \in \mathbb{R}$ , si ha

$$\int_{-\infty}^{-t} f(x) dx = \int_t^{+\infty} f(x) dx$$

(4.1.3) Sia  $X$  una variabile aleatoria assolutamente continua. Allora  $X$  è simmetrica se e solo se ammette una densità  $f$  pari, cioè tale che

$$f(x) = f(-x), \quad \forall x \in \mathbb{R}$$

Segue dalla proposizione (4.1.3) che ogni variabile aleatoria  $X$  avente legge normale standard è simmetrica. Le relazioni (4.1.1) e (4.1.2) applicate alla funzione di ripartizione  $\Phi$  danno allora rispettivamente

$$\Phi(x) + \Phi(-x) = 1$$

$$\Phi(0) = \frac{1}{2}$$

Vediamo ora in che modo si usano le tavole della legge normale standard. Esistono vari tipi di tavole. Le più comuni riportano il valore di  $\Phi(x)$  per  $x \in (0, 3)$ . Il valore  $x$  si legge fino al primo decimale nella colonna di sinistra; se  $x$  ha anche un secondo decimale, esso si legge nella riga in alto. Allora  $\Phi(x)$  si trova, nella tavola, nella posizione in cui si incrociano la riga e la colonna prescelte. Ad esempio, se  $x = 1.64$ , il valore 1.6 si legge nella colonna di sinistra. Ad esso si aggiunge il numero 0.04, che troviamo nella riga in alto. Dunque  $\Phi(x) = 0.9495$  (è il numero che si trova all'intersezione della riga marcata 1.6 con la colonna marcata 0.04).

Questo per quanto riguarda i numeri  $x \in (0, 3)$ . Se invece  $x \in (-3, 0)$  si usa la relazione (4.1.1). Ad esempio, per  $x = -1.64$  si trova

$$\Phi(-1.64) = 1 - \Phi(1.64) = 1 - 0.9495 = 0.0505$$

Infine, se  $x \geq 3$ , il valore di  $\Phi(x)$  si approssima con 1 (e, di conseguenza,  $\Phi(x) \approx 0$  per  $x \leq -3$ ).

(4.2) Siano  $X_1, X_2, \dots, X_n$   $n$  variabili aleatorie indipendenti e tutte con densità normale standard. Poniamo

$$Y = X_1^2 + X_2^2 + \dots + X_n^2$$

La densità di  $Y$  viene chiamata **densità del chi quadro a  $n$  gradi di libertà** e viene indicata con il simbolo  $\chi^2(n)$ .

(4.2.1) Calcolo della speranza di  $\chi^2(n)$ :

Per come abbiamo definito sopra la  $Y$  abbiamo

$$\mathbf{E}[Y] = \mathbf{E}[X_1^2 + X_2^2 + \dots + X_n^2]$$

per le note proprietà della speranza

$$\mathbf{E}[Y] = \mathbf{E}[X_1^2] + \mathbf{E}[X_2^2] + \dots + \mathbf{E}[X_n^2]$$

ricordando che la varianza di una variabile aleatoria standard coincide con la media del quadrato della variabile, otteniamo

$$\mathbf{E}[Y] = \mathbf{Var}X_1 + \mathbf{Var}X_2 + \dots + \mathbf{Var}X_n = n$$

(4.3) Siano  $X \sim N(0, 1)$  e  $Y \sim \chi^2(n)$  due variabili aleatorie indipendenti. Poniamo

$$Z = \sqrt{n} \frac{X}{\sqrt{Y}}$$

La densità di  $Z$  si chiama **densità  $t$  di Student a  $n$  gradi di libertà**: la indicheremo con  $t(n)$ .

Per i nostri scopi ci basterà sapere che si tratta di una densità pari e che per  $n \rightarrow \infty$  il grafico della densità  $t(n)$  tende a diventare quello della gaussiana standard.

Le tre densità  $N(0, 1)$ ,  $\chi^2(n)$  e  $t(n)$  sono collegate da un importante teorema.

(4.4) **Teorema di Cochran** – Siano  $X_1, X_2, \dots, X_n$   $n$  variabili aleatorie indipendenti, tutte con legge  $N(\mu, \sigma^2)$ . Indicheremo al solito con  $\bar{X}$  la loro media campionaria e porremo inoltre

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Consideriamo le due variabili aleatorie

$$Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

e

$$W = \frac{S^2}{\sigma^2} (n-1)$$

Allora  $Z$  ha densità  $N(0, 1)$ ,  $W$  ha densità  $\chi^2(n-1)$  ed inoltre  $Z$  e  $W$  sono tra loro indipendenti.

(4.5) **Corollario** – Nelle ipotesi e con le notazioni del teorema precedente la variabile aleatoria

$$T = \sqrt{n-1} \frac{Z}{\sqrt{W}}$$

ha densità  $t(n-1)$ . La variabile  $T$  sopra definita si può scrivere anche in una forma che risulterà utile in statistica. Si ha

$$T = \sqrt{n-1} \frac{Z}{\sqrt{W}} = \sqrt{n-1} \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \frac{\sigma}{S \sqrt{n-1}} = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

dove con  $S$  si è indicata la radice quadrata aritmetica di  $S^2$ .

## 5. La legge dei grandi numeri

Sullo spazio di probabilità  $(\Omega, A, P)$  siano  $(X_n)_{n \in \mathbb{N}}$  una successione di variabili aleatorie,  $X$  una variabile aleatoria.

(5.1) Si dice che  $(X_n)$  converge *quasi certamente* verso  $X$  se

$$(i) \quad \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} \in A$$

$$(ii) \quad P(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1$$

Vale il seguente teorema

(5.2) **Teorema (Legge forte dei grandi numeri)** – Sia  $(X_n)_{n \in \mathbb{N}}$  una successione di variabili aleatorie i.i.d., con media  $\mu$  e varianza  $\sigma^2$  finite. Allora la successione delle medie campionarie

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

converge quasi certamente a  $\mu$  per  $n \rightarrow \infty$ .

## 6. Riepilogo delle principali densità

Nome	Funzione di ripartizione	Densità	Legge
Uniforme	$F(t) = \begin{cases} 0 & \text{se } t < a \\ \frac{t-a}{b-a} & \text{se } a \leq t \leq b \\ 1 & \text{se } t > b \end{cases}$	$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$	$X \sim U(a, b)$
Esponenziale	$F(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ 1 - e^{-\lambda t} & \text{per } t > 0 \end{cases}$	$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \lambda e^{-\lambda x} & \text{se } x \geq 0 \end{cases}$	$X \sim \mathcal{E}(\lambda)$
Weibull	–	$f(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$	$X \sim W(\lambda, \alpha)$
Normale	–	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$	$X \sim N(\mu, \sigma^2)$
Normale standard	$F(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad t \in \mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$X \sim N(0, 1)$

Nome	Speranza	Varianza
Uniforme	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Esponenziale	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Weibull	–	–
Normale	$\mu$	$\sigma^2$
Normale standard	0	1

# La statistica inferenziale

## 1. Introduzione ai problemi statistici

Ogni rilevamento statistico produce un campione di dati (relativamente piccolo).

La statistica descrittiva si occupa di organizzare e riassumere in modo significativo questi dati.

La statistica inferenziale, invece, utilizzando metodi e nozioni del calcolo della probabilità, cerca di fare previsioni sul futuro, o di ottenere risultati estendibili all'intera popolazione (a partire solo dal piccolo campione osservato).

Nella pratica, tipicamente si deve effettuare un esperimento che produce una variabile aleatoria  $X$  di cui non si conosce la legge, e si vogliono ricavare informazioni su di essa.

## 2. Il concetto di stimatore

Uno dei problemi che si presentano più frequentemente allo statistico che voglia ottenere informazioni su una data variabile aleatoria  $X$ , è quello di deciderne la legge: ad esempio, può succedere che egli sappia che tale legge appartiene ad una data famiglia, dipendente da un parametro  $\theta$  non noto. Nel caso generale il parametro da studiare è indicato con  $\theta$ . In casi specifici il simbolo usato potrà essere diverso.

La cosa più naturale, che uno sperimentatore può fare, è quella di “procurarsi delle osservazioni” del fenomeno di cui  $X$  è l'espressione, effettuando qualche tipo di esperimento, e decidere (in statistica, “fare inferenza”) in base ai risultati ottenuti.

Tipicamente come risultato del suo esperimento egli otterrà dei numeri  $x_1, x_2, \dots, x_n$ . Essi vanno pensati come valori assunti (nel corso dell'esperimento) da certe variabili aleatorie  $X_1, X_2, \dots, X_n$  aventi una legge congiunta dipendente da  $\theta$ . Le chiameremo **osservazioni** di  $X$ . In corrispondenza i numeri  $x_1, x_2, \dots, x_n$  (che sono i valori assunti dalle osservazioni dopo che l'esperimento è stato effettuato) si chiameranno **valori osservati**.

Un caso molto frequente è quello in cui  $X_1, X_2, \dots, X_n$  sono tra loro indipendenti ed hanno tutte la stessa legge di  $X$ : è la formalizzazione matematica del caso in cui lo sperimentatore decide di ripetere  $n$  volte, in condizioni di indipendenza, proprio l'esperimento che produce  $X$ . Si dice allora che  $X_1, X_2, \dots, X_n$  costituiscono un **campione di taglia  $n$  estratto** dalla legge di  $X$ .

Comunque sia, lo sperimentatore userà i numeri trovati per calcolare, a partire da essi, una stima del parametro incognito; in altri termini sceglierà un'opportuna funzione  $t$  di  $n$  variabili reali e stimerà il parametro  $\theta$  con il numero  $t(x_1, x_2, \dots, x_n)$ . Ovviamente la funzione  $t$  andrà scelta non dipendente dal parametro incognito (dato che essa va usata appunto per stimarlo).

(2.1) Sia  $t: \mathbb{R}^n \rightarrow \mathbb{R}$  una funzione non dipendente da  $\theta$ .

- (i) Si chiama **stimatore** di  $\theta$  la variabile aleatoria  $T = t(X_1, X_2, \dots, X_n)$ , dove  $X_1, X_2, \dots, X_n$  sono  $n$  osservazioni di  $X$
- (ii) Si chiama **stima** di  $\theta$  il numero  $t = t(x_1, x_2, \dots, x_n)$ , dove  $x_1, x_2, \dots, x_n$  sono gli  $n$  valori osservati (corrispondenti alle osservazioni del punto (i))

(2.2) Siano  $X_1, X_2, \dots, X_n$   $n$  osservazioni aventi tutte la stessa legge, dipendente da un parametro  $\theta$ . Sia  $T = t(X_1, X_2, \dots, X_n)$  uno stimatore di una funzione  $\psi(\theta)$  del parametro. La situazione “ideale” sarebbe che valesse l’uguaglianza

$$(2.2.1) \quad T(\omega) = \psi(\theta) \quad \forall \omega \in \Omega$$

cioè che lo stimatore fornisca sempre e esattamente la quantità da stimare. Ciò non è ovviamente possibile; più ragionevole è chiedersi se l’uguaglianza (2.2.1) possa valere almeno in media; in effetti una buona proprietà di uno stimatore è la seguente

(2.3) Lo stimatore  $T = t(X_1, X_2, \dots, X_n)$  di  $\psi(\theta)$  si dice **corretto** (o **non distorto**) se vale la relazione

$$\mathbf{E}^\theta [T] = \psi(\theta) \quad \forall \theta \in \Theta$$

**Nota:** In tutti gli esempi che seguono, indicheremo con  $\mu$  e  $\sigma^2$  rispettivamente la media e la varianza della comune legge delle  $X_i$ . Nelle scritture del tipo  $\mathbf{E}^\theta$  o  $\mathbf{Var}^\theta$  o simili sottintenderemo il parametro  $\theta$ , (cioè scriveremo semplicemente  $\mathbf{E}$  o  $\mathbf{Var}$ ).

(2.3.1) La media campionaria  $\bar{X}$  è uno stimatore corretto di  $\mu$ . Infatti

$$\mathbf{E}[\bar{X}] = \mathbf{E}\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n} \mathbf{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

(2.3.2) Se  $\mu$  è nota, lo stimatore

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

è uno stimatore corretto di  $\sigma^2$ . Infatti

$$\begin{aligned} \mathbf{E}\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}\right] &= \frac{1}{n} \mathbf{E}\left[\sum_{i=1}^n (X_i - \mu)^2\right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[(X_i - \mu)^2] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{Var} X_i = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2 \end{aligned}$$

(2.3.3) Supponiamo ora in più che le osservazioni  $X_1, X_2, \dots, X_n$  siano tra loro indipendenti. Vogliamo trovare uno stimatore corretto di  $\sigma^2$  nel caso, molto frequente, che  $\mu$  non sia nota. L'idea è quella di sostituire  $\mu$  con il suo stimatore  $\bar{X}$  nella formula della (2.2.3), cioè di usare lo stimatore

$$Z = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Calcoliamo dunque  $E[Z]$ , cominciando dal numeratore della frazione.

$$\begin{aligned} (*) \quad E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= E\left[\sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X}^2\right] = E\left[\sum_{i=1}^n X_i^2 - 2n\bar{X} \frac{\sum_{i=1}^n X_i}{n} + n\bar{X}^2\right] \\ &= E\left[\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right] = E\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right] = \sum_{i=1}^n E[X_i^2] - nE[\bar{X}^2] \end{aligned}$$

D'altra parte

$$(**) \quad E[X_i^2] = \text{Var}X_i + E[X]^2 = \sigma^2 + \mu^2$$

$$\begin{aligned} (***) \quad E[\bar{X}^2] &= \text{Var}\bar{X} + E[\bar{X}]^2 = \text{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) + \mu^2 = \frac{1}{n^2} \sum_{i=1}^n \text{Var}X_i + \mu^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 + \mu^2 \\ &= \frac{1}{n^2} n\sigma^2 + \mu^2 = \frac{1}{n} \sigma^2 + \mu^2 \end{aligned}$$

Usando le relazioni (\*\*) e (\*\*\*) nella (\*) si ottiene

$$E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = \sum_{i=1}^n (\sigma^2 + \mu^2) - n\left(\frac{1}{n} \sigma^2 + \mu^2\right) = (n-1) \sigma^2$$

Dunque

$$E[Z] = \frac{n-1}{n} \sigma^2$$

che è uno stimatore distorto. Tuttavia per  $n \rightarrow \infty$ ,  $E[Z]$  converge a  $\sigma^2$  (in questo caso diremo che lo stimatore è *asintoticamente corretto* e la quantità  $\delta = \sigma^2 - E[Z] = \frac{\sigma^2}{n}$  è detta *distorsione*).

Il calcolo appena fatto per  $E[Z]$  suggerisce che lo stimatore corretto di  $\sigma^2$  è

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$



(2.4) Sostituendo il vero valore di  $\psi(\theta)$  con il suo stimatore  $T = t(X_1, X_2, \dots, X_n)$  si commette un errore, che è opportuno misurare. Si chiama **rischio quadratico** dello stimatore  $T$  la funzione  $\theta \mapsto R_T(\theta)$  definita su  $\Theta$  da

$$R_T(\theta) = \mathbf{E}^\theta[(T - \psi(\theta))^2]$$

(2.4.1) Se lo stimatore è corretto (cioè quando  $\mathbf{E}^\theta[T] = \psi(\theta)$ ), si ha evidentemente

$$R_T(\theta) = \mathbf{E}^\theta[(T - \mathbf{E}^\theta[T])^2] = \mathbf{Var}^\theta T$$

(2.4.2) In presenza di due stimatore  $S$  e  $T$  della quantità  $\psi(\theta)$ , preferiremo ovviamente lo stimatore con rischio più piccolo, per ogni valore di  $\theta$ . Si dice che  $S$  è **preferibile** a  $T$  se  $R_S(\theta) \leq R_T(\theta) \forall \theta \in \Theta$ ; se in più  $\exists \theta_0 \in \Theta$  tale che  $R_S(\theta_0) < R_T(\theta_0)$ , allora si dice che  $S$  è **strettamente preferibile** a  $T$ . Uno stimatore è detto **ammissibile** se non esistono stimatori ad esso strettamente preferibili.

(2.5) Un'altra proprietà asintotica che può essere importante per uno stimatore  $T_n = t_n(X_1, X_2, \dots, X_n)$  è la “consistenza” (o anche “coerenza”).

(2.5.1) Una successione  $(T_n)$  di stimatori di  $\psi(\theta)$  si dice **fortemente consistente** se  $T_n$  converge quasi certamente verso  $\psi(\theta)$ .

(2.5.2) Una successione  $(T_n)$  di stimatori di  $\psi(\theta)$  si dice **debolmente consistente** se  $T_n$  converge verso  $\psi(\theta)$  in probabilità (cioè,  $\lim_{n \rightarrow \infty} P(|T_n - \psi(\theta)| < \varepsilon) = 1, \forall \varepsilon > 0$ ).

(2.6) Sia  $(T_n)$  una successione di stimatori fortemente consistente del parametro  $\theta$ , e supponiamo che  $\theta \mapsto \psi(\theta)$  sia una funzione continua. Allora  $U_n \mapsto \psi(T_n)$  è una successione di stimatori fortemente consistente di  $\psi(\theta)$ .

(2.7) Sia  $X$  una variabile aleatoria la cui legge dipende da un certo numero di parametri  $\theta_1, \theta_2, \dots, \theta_r$  (il parametro  $\theta$  è in generale un vettore  $\theta = (\theta_1, \theta_2, \dots, \theta_r)$ ). Supponiamo che  $\theta_1, \theta_2, \dots, \theta_r$  siano non noti, e come al solito il nostro scopo è darne una stima dipendente dalle osservazioni  $(X_1, X_2, \dots, X_n)$ . Sia  $k$  un numero intero  $\geq 1$  fissato. Si chiama **momento teorico di ordine  $k$**  di  $X$  il numero

$$m_k = \mathbf{E}[X^k]$$

(a patto che esso sia finito, e cioè che  $\mathbf{E}[|X|^k] < \infty$ ).

(2.7.1) Dato che la legge di  $X$  dipende da  $\theta_1, \theta_2, \dots, \theta_r$ , lo stesso accadrà per il momento teorico  $m_k$ ; in altre parole esisterà una funzione  $f_k$  di  $r$  variabili tale che

$$m_k = f_k(\theta_1, \theta_2, \dots, \theta_r)$$

Supponiamo che la variabile aleatoria  $X$  ammetta i primi  $q$  momenti ( $q$  è un numero intero  $\geq 1$ ). Ciò significa che possiamo scrivere la relazione precedente  $\forall k = 1, \dots, q$ , ottenendo così il sistema

$$\begin{cases} m_1 = f_1(\theta_1, \theta_2, \dots, \theta_r) \\ m_2 = f_2(\theta_1, \theta_2, \dots, \theta_r) \\ \vdots \\ m_q = f_q(\theta_1, \theta_2, \dots, \theta_r) \end{cases}$$

Si tratta evidentemente di un sistema di  $q$  equazioni nelle  $r$  incognite  $\theta_1, \theta_2, \dots, \theta_r$  che si può cercare di risolvere. Se questo è possibile, otterremo  $r$  espressioni del tipo seguente

$$(2.7.2) \quad \begin{cases} \theta_1 = g_1(m_1, m_2, \dots, m_q) \\ \theta_2 = g_2(m_1, m_2, \dots, m_q) \\ \vdots \\ \theta_r = g_r(m_1, m_2, \dots, m_q) \end{cases}$$

(2.7.3) Supponiamo che le osservazioni  $(X_1, X_2, \dots, X_n)$  siano tra loro indipendenti. Si definisce **momento empirico** di  $X$  la quantità aleatoria

$$\hat{m}_k = \frac{\sum_{i=1}^n X_i^k}{n}$$

Per la legge dei grandi numeri, si ha

$$(2.7.4) \quad \hat{m}_k \rightarrow_{n \rightarrow \infty} m_k$$

Questa relazione spiega i nomi di momento teorico e momento empirico dati alle due quantità sopra definite; ci dice anche che i momenti empirici sono stimatori consistenti di quelli teorici.

La relazione (2.7.4) suggerisce il procedimento seguente: dato che per  $n$  grande  $\hat{m}_k \cong m_k$ , nel sistema (2.7.2) sostituiamo  $\hat{m}_k$  al posto di  $m_k$  per ogni  $k = 1, \dots, q$ , ottenendo le relazioni

$$\begin{cases} \theta_1 \cong g_1(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \\ \theta_2 \cong g_2(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \\ \vdots \\ \theta_r \cong g_r(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \end{cases}$$

ovvero avremo espresso (approssimativamente) ciascuno dei parametri in termini delle variabili aleatorie  $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q$  che sono note, perché dipendenti solo dalle osservazioni.

Dunque, a sua volta, ciascuna delle funzioni  $g_i(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q)$ , per ogni  $i = 1, \dots, r$ , dipende solo dalle osservazioni, ed è dunque uno stimatore  $\theta_i = \theta_i(X_1, X_2, \dots, X_n)$  di  $\theta_i$ .

(2.7.5) Il vettore di variabili aleatorie  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)$  si chiama **stimatore dei momenti** del parametro  $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ .

(2.7.6) Chiameremo **stime dei momenti** i numeri che si ottengono mettendo i valori osservati, cioè  $x_1, x_2, \dots, x_n$  al posto delle osservazioni  $X_1, X_2, \dots, X_n$  nelle espressioni, sopra definite,  $\theta_i(X_1, X_2, \dots, X_n)$ .

Gli stimatori dei momenti sono spesso distorti, ma consistenti. Tuttavia essi non sono dei buoni stimatori. Stimatori migliori si ottengono con il **metodo della massima verosimiglianza**.

(2.8) Sia  $X$  una variabile aleatoria la cui legge dipende da un parametro  $\theta$ . Disponiamo di  $n$  osservazioni di  $X$  (non necessariamente indipendenti) che indichiamo con  $(X_1, X_2, \dots, X_n)$ . Per semplicità, in questo momento supporremo che la legge congiunta delle osservazioni sia discreta, e indicheremo con  $P^\theta$  la loro densità congiunta. Dopo aver effettuato l'esperimento, il campione di osservazioni avrà prodotto un campione di  $n$  valori osservati. La probabilità (in funzione di  $\theta \in \Theta$ ) che il risultato sia quello effettivamente osservato è

$$\theta \mapsto P^\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n), \quad \theta \in \Theta$$

Essa è evidentemente una funzione di  $\theta$  dipendente dai valori osservati  $x_1, x_2, \dots, x_n$ . Nel caso generale essa si chiama **funzione di verosimiglianza** e il simbolo usato sarà

$$\theta \mapsto L(\theta | x_1, x_2, \dots, x_n), \quad \theta \in \Theta$$

Supponiamo di essere riusciti a calcolare (in qualche modo) il massimo di questa funzione (al variare di  $\theta \in \Theta$ ); il corrispondente punto di massimo dipenderà anch'esso dai parametri  $x_1, x_2, \dots, x_n$ ; indichiamo con  $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$ . Questo numero si chiama **stima di massima verosimiglianza** di  $\theta$ . La variabile aleatoria  $\hat{\theta}(X_1, X_2, \dots, X_n)$  si chiama **stimatore di massima verosimiglianza**.

(2.8.1) Diamo ora qualche dettaglio di alcuni metodi per il calcolo della funzione di verosimiglianza. Una situazione semplice è quella in cui il vettore delle osservazioni ha densità (congiunta) discreta. Un caso particolare di questa situazione si ha quando le osservazioni costituiscono un campione, cioè sono tra loro indipendenti. Supponiamo cioè che la variabile aleatoria  $X$  abbia densità discreta  $p_\theta(x) = P^\theta(X = x)$ . Allora

$$P^\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p_\theta(x_1) \cdot p_\theta(x_2) \cdot \dots \cdot p_\theta(x_n) = L(\theta | x_1, x_2, \dots, x_n)$$

Per analogia, se  $X$  è assolutamente continua, con densità  $f_\theta(x)$  (e naturalmente le osservazioni sono ancora tra loro indipendenti), la funzione di verosimiglianza è definita dalla formula

$$L(\theta | x_1, x_2, \dots, x_n) = f_\theta(x_1) \cdot f_\theta(x_2) \cdot \dots \cdot f_\theta(x_n)$$

Per concludere, diamo un rapido elenco di alcuni metodi usati per trovare il punto di massimo di  $L$ .

- 1) Supponiamo che l'insieme  $\Theta$  sia costituito da un numero finito di elementi  $\theta_1, \theta_2, \dots, \theta_M$ . Se  $M$  non è troppo grande, un modo del tutto elementare consiste nel calcolare  $L(\theta_i) \forall i = 1, \dots, M$  e scegliere il valore  $\theta_i$  per cui  $L(\theta_i)$  risulta massimo. Se  $M$  è grande, bisognerà ricorrere a qualche trucco, da vedere caso per caso.
- 2) Se  $\Theta$  è un intervallo della retta (eventualmente non limitato), si possono applicare i metodi studiati in Analisi I per trovare i punti di minimo e di massimo di una funzione: in particolare, la stima di massima verosimiglianza, se interna all'intervallo  $\Theta$ , è soluzione dell'equazione

$$\frac{d}{d\theta} L(\theta | x_1, x_2, \dots, x_n) = 0$$

Si ricordi che per gli (eventuali) estremi dell'intervallo, non si può applicare il criterio dei punti stazionari.

- 3) Nel caso che la stima di massima verosimiglianza sia stata ottenuta con il solo criterio dei punti stazionari, necessario ricordare che il fatto che un punto sia stazionario non garantisce automaticamente che esso sia di massimo. Di regola, bisognerebbe proseguire l'indagine studiando la derivata seconda di  $L$ .
- 4) In molti casi, prima di effettuare la derivazione di  $L$ , conviene passare al logaritmo, cioè derivare non la funzione  $\theta \mapsto L(\theta | x_1, x_2, \dots, x_n)$ , ma  $\theta \mapsto \log L(\theta | x_1, x_2, \dots, x_n)$ . Il passaggio al logaritmo non cambia i punti stazionari, né la loro natura. In questi casi, dunque, la stima di massima verosimiglianza viene trovata come soluzione dell'equazione

$$\frac{d}{d\theta} \log L(\theta | x_1, x_2, \dots, x_n) = 0$$

che è nota come *equazione di verosimiglianza*.

(2.8.2) Esempio – Calcolare la stima e lo stimatore di massima verosimiglianza del parametro  $p$  della legge bernoulliana  $B(1, p)$ , basato sul campione  $(X_1, X_2, \dots, X_n)$ . Si suppone che  $p \in (0, 1)$ . Sia  $X$  una variabile aleatoria di legge  $B(1, p)$ ; la sua densità è

$$P^p(X = x) = p^x (1 - p)^{1-x}, \quad x = 0, 1$$

Dunque, la funzione di verosimiglianza è

$$p \mapsto L(p \mid x_1, x_2, \dots, x_n) = p^{x_1} (1 - p)^{1-x_1} \cdot p^{x_2} (1 - p)^{1-x_2} \cdot \dots \cdot p^{x_n} (1 - p)^{1-x_n}$$

per  $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$ . Per semplicità poniamo

$$\alpha = x_1 + x_2 + \dots + x_n$$

L'equazione di verosimiglianza è

$$0 = \frac{d}{dp} \log L(p \mid x_1, x_2, \dots, x_n) = \frac{d}{dp} \log(p^\alpha (1 - p)^{n-\alpha}) = \frac{d}{dp} (\alpha \log p + (n - \alpha) \log(1 - p)) = \frac{\alpha}{p} - \frac{n - \alpha}{1 - p}$$

la cui unica soluzione è

$$\hat{p} = \frac{\alpha}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}$$

(media del campione di valori osservati). Lo stimatore di massima verosimiglianza è dunque

$$\hat{p} = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}$$

ovvero la media campionaria delle osservazioni.

### 3. Intervalli di fiducia

Ora ogni funzione delle osservazioni, non dipendente dal parametro incognito  $\theta$ , sarà detta **statistica**. Dunque il concetto di statistica non è diverso da quello di **stimatore**, che abbiamo dato in precedenza. Quello che cambia è solo il nome: ora non si tratta di dare un valore approssimato al parametro, ma di utilizzare la nostra funzione delle osservazioni per confronti più specifici.

(3.1) Sia  $\alpha \in (0, 1)$  un numero fissato. Date due statistiche  $T_1 = t_1(X_1, X_2, \dots, X_n)$  e  $T_2 = t_2(X_1, X_2, \dots, X_n)$ , si dice che  $I_X = [T_1, T_2]$  è un **intervallo di fiducia per  $\psi(\theta)$  di livello  $1 - \alpha$**  se,  $\forall \theta \in \Theta$  si ha

$$P^\theta(\psi(\theta) \in I_X) \geq 1 - \alpha$$

(3.1.1) Tipicamente il valore di  $\alpha$  è piccolo ( $\alpha = 0.05; \alpha = 0.01 \dots$ ).

(3.1.2) Il significato della definizione (3.1) è il seguente: in base alle osservazioni che abbiamo, possiamo dire che  $T_1 \leq \psi(\theta) \leq T_2$  con probabilità almeno  $1 - \alpha$ , e questo  $\forall \theta$ .

(3.1.3) Costruiamo un intervallo di fiducia di livello 0.95 per il parametro  $\lambda$  dell'esponenziale, basato su una sola osservazione  $X$ . Significa che si devono trovare due funzioni  $t_1(X)$  e  $t_2(X)$  tali che

$$P^\lambda(t_1(X) \leq \lambda \leq t_2(X)) \geq 0.95$$

Partiamo da questa semplice osservazione: se  $X \sim \text{Esp}(\lambda)$  (con media pari a  $1/\lambda$ ), allora la variabile  $Y = \lambda X \sim \text{Esp}(1)$ . Infatti, per  $t > 0$  si ha

$$P(Y \leq t) = P(\lambda X \leq t) = P(X \leq t/\lambda) = F_X(t/\lambda) = 1 - e^{-t}$$

Di conseguenza,  $\forall a, b > 0$  con  $a < b$  si ottiene

$$P^\lambda(a \leq Y \leq b) = (1 - e^{-b}) - (1 - e^{-a}) = e^{-a} - e^{-b}$$

il che equivale a

$$P^\lambda(a \leq \lambda X \leq b) = P^\lambda\left(\frac{a}{X} \leq \lambda \leq \frac{b}{X}\right) = e^{-a} - e^{-b}$$

Allora poniamo

$$t_1(X) = \frac{a}{X}, \quad t_2(X) = \frac{b}{X}$$

dove le costanti  $a$  e  $b$  sono scelte in modo che  $e^{-a} - e^{-b} = 0.95$ .

Il metodo qui seguito è quello della **quantità pivotale**.

**Nota:** D'ora in avanti, gli intervalli che costruiremo saranno tutti di livello  $1 - \alpha$  fissato. Inoltre, ai punti (3.2), (3.3), (3.4) e (3.5)  $(X_1, X_2, \dots, X_n)$  sarà un campione di taglia  $n$  estratto dalla legge  $N(\mu, \sigma^2)$ .

(3.2) **Intervalli di fiducia per la media della normale con varianza nota** – Si parte dall'osservazione che

$$Y = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1)$$

(a) **Intervallo bilaterale** – Sia

$$(*) \quad P^\mu(a \leq Y \leq b) = \Phi(b) - \Phi(a)$$

Pertanto ora basterà trovare  $a$  e  $b$  in modo che  $\Phi(b) - \Phi(a) = 1 - \alpha$ .

Siano ora  $\beta$  e  $\gamma$  due numeri reali  $\in [0, 1]$  tali che  $b = \phi_\beta$  e  $a = \phi_\gamma$  (si pensi a  $\phi$  come la funzione inversa di  $\Phi$ ); allora si ha

$$\Phi(b) = \beta, \quad \Phi(a) = \gamma$$

Pertanto basterà scegliere  $\beta$  e  $\gamma$  in modo tale che  $\beta - \gamma = 1 - \alpha$ . Una scelta possibile è  $\beta = 1 - \alpha/2$ ,  $\gamma = \alpha/2$ . Dalla (\*) si ricava

$$\begin{aligned} \Phi(b) - \Phi(a) &= P^\mu(a \leq Y \leq b) = P^\mu\left(a \leq \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq b\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} b \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} a\right) \\ &= P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_\beta \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_\gamma\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{\alpha/2}\right) = \\ &P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2}\right) \end{aligned}$$

L'intervallo è dunque

$$\left( \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \right)$$

(b) **Intervallo unilaterale destro** – Il termine significa che si vuole trovare una limitazione per  $\mu$  solo dal basso. Questa volta partiamo dalla relazione

$$(**) \quad P^\mu(Y \leq b) = \Phi(b)$$

Sia ora  $\beta$  come al punto (a). Pertanto basterà trovare  $b$  in modo che  $\Phi(b) = 1 - \alpha = \beta$ . Ovvero, semplicemente,  $b = \phi_{1-\alpha}$ . Dalla (\*\*) si ricava che

$$\begin{aligned} \Phi(b) &= P^\mu(Y \leq b) = P^\mu\left(\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq b\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} b \leq \mu\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_\beta \leq \mu\right) \\ &= P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha} \leq \mu\right) \end{aligned}$$

L'intervallo è dunque

$$\left( \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro** – Analogamente al caso dell'intervallo unilaterale destro troviamo che l'intervallo è

$$\left(-\infty, \bar{X} - \frac{\sigma}{\sqrt{n}}\phi_{\alpha}\right) = \left(-\infty, \bar{X} + \frac{\sigma}{\sqrt{n}}\phi_{1-\alpha}\right)$$

ricordando che  $-\phi_{\alpha} = \phi_{1-\alpha}$ .

(3.3) Intervalli di fiducia per la media della normale con varianza non nota – Nella pratica gli intervalli di (3.2) sono di scarsa utilità, perché nelle formule che li definiscono interviene la varianza  $\sigma^2$ , che in genere non si conosce. In questo caso si può sostituire  $\sigma^2$  con

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

(che ne è uno stimatore corretto) e applicare di nuovo il metodo della quantità pivotale partendo dalla variabile aleatoria

$$Z = \sqrt{n} \frac{\bar{X} - \mu}{S} \sim t(n-1)$$

Osservando che l'unica proprietà della legge normale standard che abbiamo usato è stata la simmetria e ricordando che anche la legge  $t$  di Student è simmetrica, risulta chiaro che tutto ciò che abbiamo detto in (3.2) si può ripetere, semplicemente sostituendo  $\sigma$  con  $S$  e i quantili della normale standard con quelli della  $t$  di Student. Pertanto avremo

(a) **Intervallo bilaterale**

$$\left(\bar{X} - \frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1), \bar{X} + \frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1)\right)$$

(b) **Intervallo unilaterale destro**

$$\left(\bar{X} - \frac{S}{\sqrt{n}}t_{1-\alpha}(n-1), +\infty\right)$$

(c) **Intervallo unilaterale sinistro**

$$\left(-\infty, \bar{X} + \frac{S}{\sqrt{n}}t_{1-\alpha}(n-1)\right)$$



(3.4) **Intervalli di fiducia per la varianza della normale con media nota** – Qui si parte osservando che la variabile aleatoria

$$Y = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

ha legge  $\chi^2(n)$ . Posto

$$U^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

(che è uno stimatore corretto della varianza) si può scrivere

$$Y = \frac{nU^2}{\sigma^2}$$

(a) **Intervallo bilaterale** – Indichiamo con  $F_n$  la funzione di ripartizione della  $\chi^2(n)$ . Allora si ha

$$F_n(b) - F_n(a) = P^{\sigma^2}(a \leq Y \leq b) = P^{\sigma^2}\left(a \leq \frac{nU^2}{\sigma^2} \leq b\right) = P^{\sigma^2}\left(\frac{nU^2}{b} \leq \sigma^2 \leq \frac{nU^2}{a}\right)$$

se, al solito come visto in (3.2), poniamo  $b = \chi_{\beta}^2(n)$ ,  $a = \chi_{\gamma}^2(n)$ , avremo

$$1 - \alpha = F_n(b) - F_n(a) = \beta - \gamma$$

Una scelta possibile (come in 3.2) è  $\beta = 1 - \alpha/2$ ,  $\gamma = \alpha/2$  e si ottiene l'intervallo

$$\left( \frac{nU^2}{\chi_{1-\alpha/2}^2(n)}, \frac{nU^2}{\chi_{\alpha/2}^2(n)} \right)$$

(b) **Intervallo unilaterale destro** – Esattamente come per la (3.2) abbiamo che l'intervallo è

$$\left( \frac{nU^2}{\chi_{1-\alpha}^2(n)}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro** – Analogamente abbiamo che l'intervallo è

$$\left( -\infty, \frac{nU^2}{\chi_{\alpha}^2(n)} \right)$$

(3.5) **Intervalli di fiducia per la varianza della normale con media non nota** – Poiché la media non è nota, useremo, al posto di  $\mu$ , il suo stimatore (corretto)  $\bar{X}$ . Prendiamo in considerazione ora la quantità

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

(che, come è noto, è uno stimatore corretto della varianza). Quindi, applichiamo il metodo della quantità pivotale a partire dalla variabile aleatoria

$$W = \frac{S^2}{\sigma^2}(n-1)$$

che, dal teorema di Cochran, risulta avere legge  $\chi^2(n-1)$  (quindi i ragionamenti fatti al punto (3.4) valgono anche qui). Dunque, per ricavare i tre nuovi intervalli basterà sostituire, nelle formule di (3.4),  $n-1$  al posto di  $n$  (e, naturalmente, i quantili della  $\chi^2(n-1)$  al posto di quelli della  $\chi^2(n)$ ). Si ottiene cioè

(a) **Intervallo bilaterale**

$$\left( \frac{(n-1)U^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)U^2}{\chi^2_{\alpha/2}(n-1)} \right)$$

(b) **Intervallo unilaterale destro**

$$\left( \frac{(n-1)U^2}{\chi^2_{1-\alpha}(n-1)}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro**

$$\left( -\infty, \frac{(n-1)U^2}{\chi^2_{\alpha}(n-1)} \right)$$

(3.6) **Intervalli di fiducia per il parametro della bernoulliana** – Sia  $(X_1, X_2, \dots, X_n)$  un campione dalla legge  $B(1, p)$ . Osservando che per  $n$  abbastanza grande, la variabile aleatoria

$$Y = \frac{X - p}{\sqrt{p(1-p)}} \sqrt{n}$$

(dove  $p$  è la media e  $p(1-p)$  è la varianza della variabile aleatoria bernoulliana) è approssimativamente di legge  $N(0,1)$  per il Teorema Limite Centrale, possiamo pensare di usufruire di quanto detto a proposito di campioni gaussiani. Per la stima bilaterale, avremo l'intervallo

$$\left( \bar{X} - \frac{\sqrt{p(1-p)}}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sqrt{p(1-p)}}{\sqrt{n}} \phi_{1-\alpha/2} \right)$$

Purtroppo, l'intervallo risultante dipende dal parametro  $p$  che è appunto l'incognita da stimare, e dunque non è un buon intervallo di fiducia. L'idea è allora quella di sostituire  $p$  con il suo stimatore  $\bar{X}$ . E si ha

(a) **Intervallo bilaterale**

$$\left( \bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha/2} \right)$$

(b) **Intervallo unilaterale destro**

$$\left( \bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro**

$$\left( -\infty, \bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha} \right)$$